Total Synthesis

DOI: 10.1002/anie.201301666

The Winding Pathway to Erythropoietin Along the Chemistry–Biology Frontier: A Success At Last**

*Rebecca M. Wilson, Suwei Dong, Ping Wang, and Samuel J. Danishefsky**

amino acids · erythropoietin · glycoprotein · synthetic methods · total synthesis

The total synthesis of a homogeneous erythropoietin (EPO), possessing the native amino acid sequence and chitobiose glycans at each of the three wild-type sites of N glycosylation, has been accomplished in our laboratory. We provide herein an account of our decade-long research effort en route to this formidable target compound. The optimization of the synergy of the two bedrock sciences we now call biology and chemistry was central to the success of the synthesis of EPO.

1. Introduction

Our laboratory recently disclosed the first total synthesis of a homogeneous, wild-type erythropoietin (EPO) glycoprotein. In this retrospective, we recount highlights of the decade-long research effort which culminated in reaching this elusive goal. In so doing, we attempt to convey the sense of the struggle in the context of a larger and broader question, that is, how to interface the capabilities of chemical synthesis with the seemingly forbidding challenges of synthesizing large biomolecules in a cell-free context without the benefit of enzymatic and nucleic acid-mediated orchestration. We viewed the EPO synthesis challenge as an opportunity to test our privately held perception that the field of chemical synthesis had matured to the point that even a glycoprotein biologic of the scope and complexity of EPO could be seen as a not totally implausible target for total synthesis. We were certainly not confident that all the pieces of the looming puzzle were already at hand. The venture was undertaken in the context of a perhaps naïve hope that the required capabilities might be filled in by drawing from the very powerful lessons of mechanistic organic chemistry. Needless to say, the field of protein synthesis has witnessed numerous impressive advances in the preparation of biologically active glycoproteins and glycopeptides.[1–6] However, as will be shown, the highly complex, multiply gycosylated EPO glycoprotein was seen to present a singularly formidable synthetic challenge.

Before launching into the EPO project per se, it is perhaps appropriate to air some general thoughts about interactions between chemistry and biology. Indeed, it is increasingly evident that many of the most inviting and challenging problems in drug discovery are only approachable by combining various scientific disciplines in a synergistic manner. In attempting to create such frontier-level scientific interdisciplinary collaborations, it soon became apparent that the diverse cultural patterns of the core discipline may give rise to differing linguistic nuances. These subtleties must be appreciated for optimal cross-disciplinary communication.

For instance, in the case of the biology-chemistry frontier of greatest interest to us, much of the cultural variations are rooted in the differing historic missions of the hardcore disciplines. Thus, biology has traditionally excelled at delineating function in extraordinary detail, often in the absence of clearly defined structures, or even homogeneous substrates. By contrast, chemistry, which tended to focus on relatively narrow, but well definable issues of structure, often (unfortunately!) overlooked momentous questions of function.

Again, in the context of their historic preoccupation with precise recognition of function versus precise definitions of structure, the term mechanism calls forth different connotations. Given the higher order of complexity of the problems traditionally encountered in biology, mechanism is seen as defining the minimum number of biocomponents necessary to bring about a functional consequence. Successful demonstration of the functional phenomenon, even in a cell-free setting, suggests that the mechanism is well in hand. By contrast,

[*] R. M. Wilson, Dr. S. Dong, Dr. P. Wang, Prof. S. J. Danishefsky
Laboratory for Bioorganic Chemistry
Sloan-Kettering Institute for Cancer Research
1275 York Avenue, New York, NY 10065 (USA)
Prof. S. J. Danishefsky
Department of Chemistry, Columbia University
Havemeyer Hall, 3000 Broadway, New York, NY 10027 (USA)
E-mail: s-danishefsky@ski.mskcc.org

[**] Dedicated to Stephen Kent and his associates for their discovery of native chemical ligation.

given the more confinable nature of the typical problem in chemistry, mechanism carries with it a higher order of detail. In chemistry, mechanistic command suggests a reasonably precise insight as to how substrates (and reagents) interact to bring about structural changes at the covalent or noncovalent levels. In principle, at least (if not always in practice), one would aspire to know how the structural features in the environs of the cleavage and bond-forming sites modulate the chemical outcome.

An indication (though not fully rigorous) as to whether one is on the right track with respect to mechanisms in chemistry is the ability to build on the insights gathered and use them to successfully craft a new chemical transformation. Thus, in chemical synthesis, as well as biological synthesis, there is an ongoing synergistic interactivity with mechanistic understanding.

The meanings of the term synthesis in biology and chemistry are often quite different in perception and scope. Generally, synthesis, as understood in biology, involves a fairly limited terrain of targets which are assembled by enzymatic means under overall cellular management. More often than not, the targets are structures of repetitive motifs (polypeptides, oligosaccharides, or oligonucleotides). Particularly in the case of polypeptides and oligonucleotides, these repeating units are assembled under some form of template management with truly impressive quality control, and (in the view of the chemist) enviable catalyst turnover and yield.

In contrast, synthesis in the language of chemistry, carries with it a broader connotation since, at least in principle, any structure capable of existence can be assembled in the laboratory. Newly developed methodologies enable fashioning of wholly new strategies with which to deal with novel targets as they come along. The concept that the only limitations to the power of synthesis to construct novel, and even highly complex targets, arise from limitations in imagination and persistence, is in our view a particularly unique message which chemistry offers to the world of "gedanken."

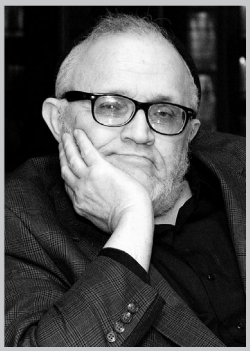
In this Minireview, we provide a particular case history wherein we undertook to synthesize EPO in our laboratory. Being chemists by trade and culture, our target was at the structural level, non-negotiable. It had to encompass the exact primary structure corresponding to the natural glycoprotein. While relevant model targets would be pursued as part of the exercise, in the end there could be no tampering with the highly conserved polypeptide structure corresponding to the real EPO. Moreover, all of the highly conserved sites of glycosidation would have to carry glycan domains joined through natural linkage modes.

As for the glycosidic domains themselves, here the problem becomes even more complicated because of the serious inhomogeneity in EPO itself. It remained to be determined how we would deal with a target which, while definable in terms of primary protein structure and biological function, represents in chemical terms a conceptual average of hypothetical individual structures, none of which was previously characterized in pure form. Given the uncharacteristically lax management of nature (for instance Chinese hamster ovary [CHO] cells) and lax control in exporting EPO, it would be left to chemistry to produce a structurally defined chemical entity and to ensure that such an EPO would meet its functional mission (i.e. erythropoiesis). It was with this charge that we proceeded to EPO.

Rebecca M. Wilson received her undergraduate degree in biochemistry at Tufts University. She went on to the University of California at Berkeley and then to Caltech, where she conducted research with Prof. David MacMillan and earned her masters degree. She then worked in the small-molecule project group at Amgen Pharmaceuticals. In 2004, she joined Prof. Samuel Danishefsky at Memorial Sloan-Kettering Cancer Center, where she is involved in various consultative and editorial undertakings.

Ping Wang received his Ph.D. at the Shanghai Institute of Organic Chemistry with Prof. Biao Yu in 2005. After postdoctoral study with Prof. Peter H. Seeberger at Sanford-Burnham Medical Research Institute (2005–2008), he joined the research group of Samuel J. Danishefsky at Memorial Sloan-Kettering Cancer Center. His research focuses on the synthesis of glycoproteins.

Suwei Dong was born in Guizhou province, People's Republic of China, in 1981. He received a B.S. degree from Peking University in 2005 and then moved to Boston University, where he focused on the total synthesis of natural products with Prof. John A. Porco. After receiving his Ph.D. in 2010, he joined Memorial Sloan-Kettering Cancer Center as a postdoctoral scholar under the mentorship of Prof. Samuel J. Danishefsky, where he works on the chemical synthesis of therapeutic polypeptides and glycoproteins including erythropoietin.



Samuel J. Danishefsky received his B.S. degree at Yeshiva University and performed graduate research with Prof. Peter Yates. He worked with Prof. Gilbert Stork at Columbia University as an NIH Postdoctoral Associate, and joined the University of Pittsburgh as an Assistant Professor in 1963. In 1980 he moved to Yale University, and in 1993 he moved to New York, where he is currently Centennial Professor of Chemistry at Columbia University and the Eugene Kettering Chair and Head of the Laboratory of Bioorganic Chemistry at Memorial Sloan-Kettering Cancer Center.

1.1. EPO: An Essential Glycoprotein Hormone

In this narrative, we will be emphasizing the technical challenges of preparing a glycoprotein as complex and challenging as EPO. Nonetheless, it is perhaps appropriate to dwell briefly on the extraordinary centrality of EPO to life itself. EPO, a 166 amino acid glycoprotein containing four sites of glycosylation (Figure 1), mediates a range of mechanistically distinct and crucial biological processes, from erythropoiesis (the production of red blood cells) to post-injury cytoprotection.[7] The critical role of EPO in promoting erythropoiesis is well understood. Onset of hypoxia (deficiency of oxygen) stimulates the production of EPO in the adult kidneys. Upon release into circulation, EPO binds to the homodimeric EPO receptor (EPOR) at low concentrations, thus initiating a cascade sequence which culminates in the inhibition of apoptosis of red blood cell precursors and, consequently, enhanced erythropoiesis. To achieve effective erythropoiesis, low levels of EPO must be maintained over a prolonged period of time. Recombinant human EPO (rhEPO) and analogues with enhanced plasma stability, are widely prescribed for the treatment of EPO-deficient anemia.

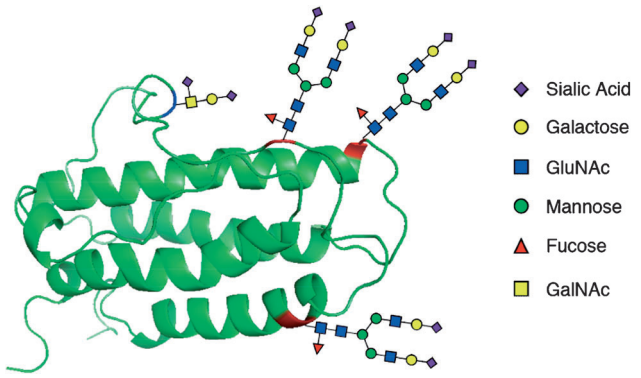


Figure 1. Ribbon structure of EPO containing a consensus sequence of N- and O-linked carbohydrate domains.

Endogenous and recombinant forms of EPO are produced as complex and inseparable mixtures of glycoforms. While it is known that individual EPO glycoforms possess different levels of biological activity and biostability,[8] it has thus far been very difficult to rigorously analyze the relative activities of individual glycoforms without access to homogeneous, single-glycoform material.[9] Despite the fact that currently available recombinant techniques are unable to deliver homogeneous samples of the glycoprotein, we (and others) well recognize that the burgeoning field of synthetic protein chemistry may offer a powerful and versatile solution to the longstanding goal of acquiring single-glycoform EPO.

Indeed, on the basis of its historical biomedical import and its complex structure, the EPO glycoprotein has emerged as an orienting benchmark target for practitioners of protein, and particularly glycoprotein, chemical synthesis. Over the past decade, a number of prominent research teams, operating at the forefront of this field, have undertaken efforts directed toward the total synthesis of homogeneous EPO. These endeavors have served both to inspire the development of groundbreaking synthetic methodologies and to highlight the significant limitations of contemporary state-of-the-art capabilities in protein synthesis.

1.2. EPO: A Benchmark Target in Biologics Synthesis

Before we commence with a personal accounting of the progression of our long-term EPO synthesis program, and in the context of the cultural issues identified above, it is well to re-emphasize that the EPO challenge has caught the interest of many researchers who have provided impressive contributions in this arena. Several groups have reached synthetic or semisynthetic EPO analogues bearing modifications within the native peptide sequence or glycosylation sites. A number of these constructs have exhibited impressive bioactivity. We provide below a brief synopsis of some pioneering research advances in the EPO field that have arisen from other synthesis-directed laboratories over the past 12 years.

1.2.1. EPO as a Synthetic Target: Previous Studies

In 2001, Flitsch and co-workers employed semisynthetic methods to gain access to homogeneous EPO glycoforms bearing glycans appended to non-native cysteine thiol groups.[10] In this study, recombinantly produced EPO mutants incorporating Asn→Cys mutations were subjected to chemical glycosylation with glycosyl-β-*N*-iodoacetamide. This early success, albeit using recombinant protein, laid the foundation for future synthetic efforts directed toward EPO. Shortly thereafter, Kent and co-workers prepared a stable and potent synthetic EPO aglycone analogue with a modified amino acid structure.[11] Installation of two polymer domains on the protein served to confer significantly enhanced in vivo circulation time and biological activity. The landmark effort by Kent et al. highlighted the power of chemical synthesis and rational design in the development of improved biologic-based lead therapeutic agents. Continuing their ongoing interest in the EPO synthesis problem, the group of Kent recently reported the synthesis of [$Lys^{24,38,83}$]EPO aglycone, wherein each of the three native N-glycosylation Asn residues had been mutated to Lys.[12] The synthetic protein was found to possess in vitro activity. The groups of Kajihara and Macmillan have also registered impressive advances in the EPO field. Richardson and Macmillan described the preparation of a semisynthetic EPO with several amino acid mutations and two glycans at both natural and unnatural glycosylation sites.[13] This approach is particularly amenable to SAR evaluations. Working together, the groups of Kajihara and Macmillan merged the tools of chemical synthesis and bacterial expression to generate an EPO analogue bearing several sites of amino acid mutation.[14] Notably, this EPO construct incorporated two 11-mer sialyloligosaccharides, which bear structural similarity to the natural EPO N-glycans. Finally, as described in a recent disclosure, the Kajihara and co-workers elegantly accomplished the preparation of a fully synthetic EPO analogue bearing an 11-mer at one of the native glycosylation sites (Asn83).[15] A summary of these

Table 1: Milestones en route to synthetic EPO.

Flitsch, 2001.
● Semi-synthetic EPO analog.
● One glycan at a mutant Cys residue.
Kent, 2003.
● Fully synthetic EPO analog.
● Non-native amino acid sequence.
● No carbohydrate groups.
● Polymer domains → potent in vivo activity.
Macmillan, 2008.
● Semi-synthetic EPO analog.
● No carbohydrate groups.
Kajihara and Macmillan, 2009.
● Semi-synthetic EPO analog.
● Non-native amino acid sequence.
● Two carbohydrate groups (11-mer glycan).
Kent, 2012.
● Fully synthetic EPO analog.
● Non-native amino acid sequence.
● No carbohydrate groups.
● Possesses in vitro activity.
Kajihara, 2012.
● Fully synthetic EPO analog
● Non-native amino acid sequence.
● One carbohydrate group (11-mer glycan).
Danishefsky, 2012.
√ Fully synthetic EPO.
√ Native amino acid sequence.
√ Carbohydrates at all native sites (chitobiose).
√ Possesses in vitro activity.

milestones en route to synthetic EPO is encapsulated in Table 1.

The preceeding discussion illuminates the extent to which EPO has served as an orienting benchmark target for chemical synthesis. Using the tools of recombinant synthesis, site-specific amino acid mutation, and chemical synthesis, researchers in this area have made impressive advances toward our understanding of EPO and the broader field of glycoprotein synthesis. However, there can be no doubt that prior to our 2012 disclosure,[16] no total synthesis of EPO had been accomplished. According to our previously defined criteria, an EPO total synthesis must deliver the glycoprotein incorporating the native amino acid sequence and carbohydrate sectors at all of the native sites of glycosylation. In 2012, our laboratory achieved the first homogeneous, fully synthetic, wild-type EPO bearing glycosylation at each of the four native sites. Moreover, our homogeneous EPO glycoprotein was folded and found to exhibit erythropoietic activity in in vitro settings.

1.2.2. Introduction to the Danishefsky EPO Program

The first person to tell us about the magic of EPO and to prod us, if only in a jocular fashion, was Richard Lerner of the Scripps Research Institute. In time, this fateful conversation would propel our laboratory into an exciting new direction, thus significantly influencing our research priorities over the next ten years and inspiring the development of a diverse menu of broadly impactful methodological innovations in the field of protein and glycoprotein synthesis.[17] Moreover, these forays in glycoprotein chemical synthesis have led us to pursue a range of therapeutically useful glycoprotein and protein targets, including parathyroid hormone (PTH),[18] parathyroid hormone-related protein (PTHrP),[19] and follicle stimulating hormone (FSH).[20]

At the outset of our EPO total synthesis program, we became aware, often painfully, of significant lacunae in the methodologies available for the synthesis of multiply glycosylated proteins. Before we could even begin to mount a serious effort toward homogeneous EPO, we would first need to bring to fruition a range of fundamental synthetic capabilities. In so doing, we well recognized that the technologies developed in response to the EPO challenge would surely accrue to the benefit of the broader field of glycoprotein and protein synthesis. The opportunity to achieve meaningful advances in the field of protein synthesis held particular appeal because a foundational mission of our laboratory has been that of devising innovative and widely useful new synthetic capabilities in the face of challenging target molecules. Indeed, the methods described below were developed in the context of the EPO synthesis program and have already been widely employed by our laboratory and others, en route to a range of therapeutically relevant biologic targets.

Specifically, the synthesis of the multiply glycosylated EPO protein target would require the development of capabilities in three key areas. First, there had to be devised methods for the convergent synthesis of complex carbohydrate domains.[21] Second, the carbohydrate and peptide domains must be efficiently merged to generate glycopeptide fragments. Third, the individual glycopeptide sectors must be joined to forge the glycoprotein primary structure. Here, we were particularly concerned with the development of methods for the reiterative coupling of multiple glycopeptide fragments and for the coupling of glycopeptides at a range of amino acid sites.

Finally, we would seek to apply these methodological advances to the synthesis of EPO, a glycoprotein of unprecedented size and complexity. This central challenge would encompass a significant degree of stress testing of our newly developed methods. These broad research areas were to be explored in parallel over the next ten years. We provide below a roughly chronological accounting of our diverse research activities which ultimately culminated in the first total synthesis of wild-type EPO. Descriptions of key methodological advances and their applications to the EPO problem are interspersed throughout this retrospective. The structures of the various oligosaccharides employed herein are depicted in Figure 2.

2. Development of Methods for the Synthesis of Glycoproteins

2.1. A Method for the Synthesis of Complex Glycopeptides from Carbohydrate and Peptide Precursors

At the outset of the program a number of methods existed for the assembly of small glycopeptides. However, our goal to

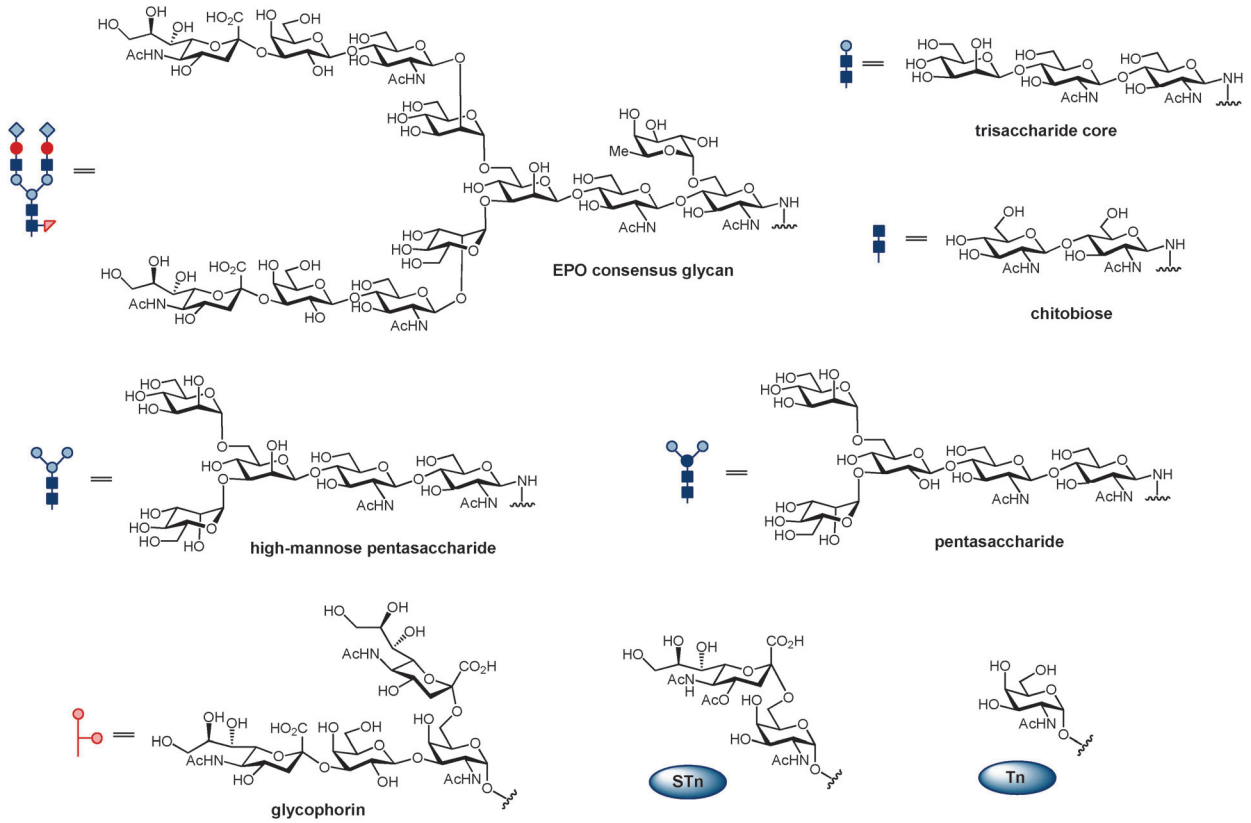
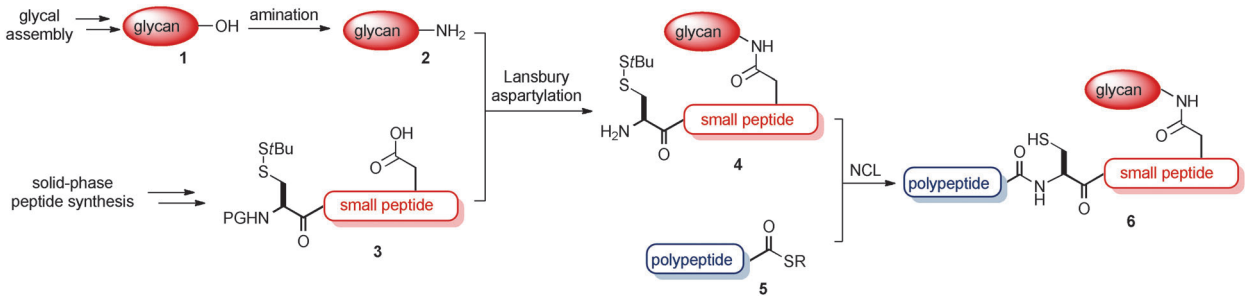


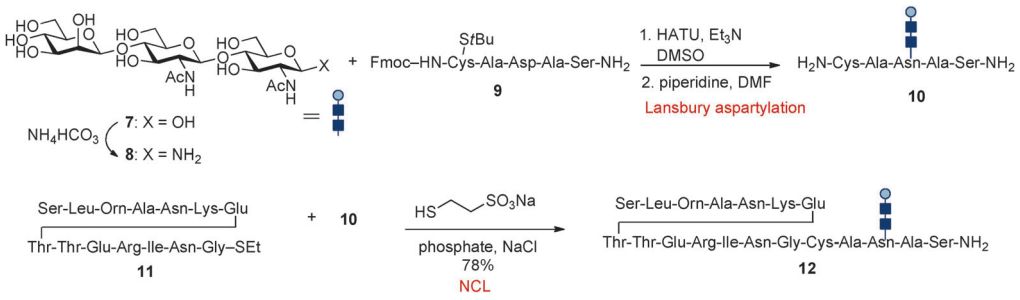
Figure 2. Structures of oligosaccharides employed in this Minireview.

establish fully synthetic routes to complex glycoproteins would require mounting large glycodomains onto extended polypeptide scaffolds. Toward this end, we first had to develop an efficient protocol by which to assemble small glycopeptides, of the type **4** (Scheme 1), from mature oligosaccharide (**2**) and peptide domains (**3**).[22] The unprotected oligosaccharide **1**—accessed through glycal assembly strategies developed in our laboratory—undergoes anomeric β-amination under Kochetkov conditions[23] to produce the glycosylamine **2**. Under the aspartylation protocol of Anisfeld and Lansbury,[24] **2** is joined to the aspartate residue of **3** to yield a short glycopeptide fragment. This intermediate is then coupled with a longer peptide domain through native chemical ligation (NCL)[25] to deliver the target glycopeptide adduct **6**. Clearly, an optimally convergent route to glycopeptides such as **6** would involve the direct coupling of the glycan domain to the full polypeptide fragment. However, severe complications arising from attempts to achieve such large fragment aspartylation precluded the adoption of such a strategy. As will be described (Section 4.2), it was only in 2012 that we mastered the complexities associated with direct aspartylation of large glycan domains with substantial polypeptides. However, prior to that discovery, the general strategy depicted in Scheme 1 was routinely adopted for the synthesis of glycopolypeptide fragments.

A demonstration of the sequential aspartylation/NCL approach is outlined in Scheme 2.[22] The fully deprotected trisaccharide **8** was coupled with the pentapeptide **9** under Lansbury conditions. After deprotection, the resultant glycopeptide **10** readily underwent NCL with the polypeptide thioester **11** to afford the fully unprotected glycopeptide adduct **12**. Importantly, a single stereoisomeric adduct was isolated and found to correspond to the β-linked glycosylamide.



Scheme 1. Glycopeptide–peptide ligation strategy. PG = protecting group.



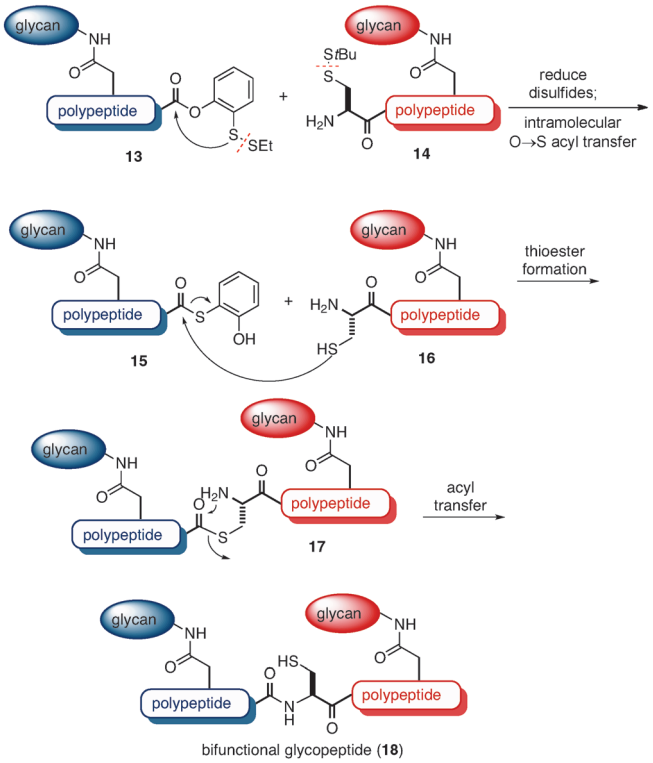
Scheme 2. Demonstration of a novel glycopeptide–peptide ligation protocol. DMSO = dimethylsulfpxide, HATU = *O*-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate.

2.2. A Method for the Synthesis of Multiply Glycosylated Peptide Domains

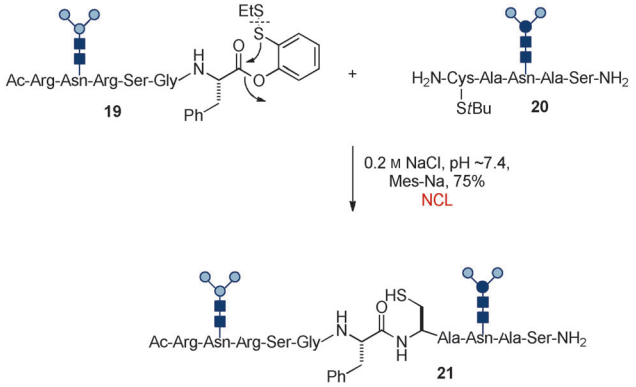
Having thus achieved the capacity to synthesize complex glycopeptide fragments bearing single glycan domains, our next challenge would be that of assembling multiply glycosylated peptides through ligation of two synthetic glycopeptide fragments. Of course, in thinking about the purely chemically mediated ligation of peptides, one is naturally drawn to consider the historic method of NCL pioneered by Kent and co-workers.[25] However, direct extension of the Kent NCL methodology employed in the peptide–glycopeptide coupling protocol described above (Scheme 1) was not considered feasible because of the difficulties inherent in synthesizing a preformed glycopeptide thioester. In 2004, our laboratory described the first glycopeptide–glycopeptide ligation protocol, which featured our novel *o*-mercaptoaryl ester rearrangement (OMER) technology. The general concept is shown in Scheme 3. An inert *ortho*-thiophenolic ester is installed on the C-terminal glycopeptide (**13**).[26] Upon disulfide reduction, the phenolic moiety undergoes O→S migration, thus providing an in situ thioester (**15**) which is activated for intermolecular thioester exchange with the cysteine residue of the glycopeptide coupling partner (**16**). Following spontaneous intramolecular S→N acyl transfer (**17**→**18**), the bi-domainal glycopeptide adduct is in hand. This approach is now used extensively by our laboratory and others as an efficient means by which to achieve ligations in systems where the C-terminal coupling fragment bears a sensitive carbohydrate domain. Moreover, this OMER methodology has also been successfully applied to the syntheses of cyclic peptide and glycopeptide systems.[27]

At the outset, we had postulated the existence of an unfavorable but dynamic equilibrium, whereby only small quantities of a highly reactive thioester species (**15**) would be formed. However, in a key mechanistic study[28] it was subsequently observed that the thioester is in fact the predominant species at the low pH levels used in these ligations. As will be seen (see Section 2.3), this mechanistic observation has been exploited for the development of noncysteine based NCL variants.

A demonstration of the synthetic potential of the OMER methodology is depicted in Scheme 4.[26] The glycopeptide precursors **19** and **20** were prepared through standard synthetic methods. After disulfide reduction, the substrates



Scheme 3. Glycopeptide–glycopeptide ligation strategy.



Scheme 4. Demonstration of a glycopeptide–glycopeptide ligation protocol.

underwent the hoped-for cysteine-mediated ligation to deliver the bifunctional glycopeptide **21** bearing two differentiated carbohydrate sectors.

2.2.1. Application to EPO: Synthesis of the EPO(22–37) Glycopeptide Fragment

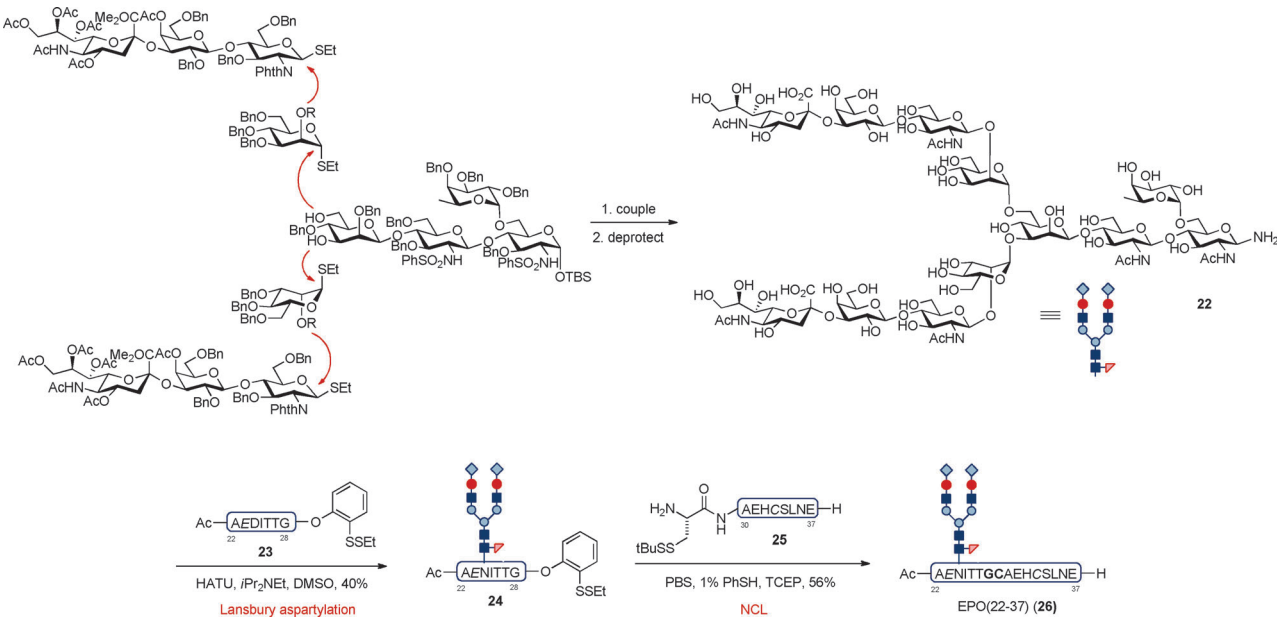
Having developed efficient methods for the synthesis of complex glycopeptides, we next sought to accomplish a synthesis of the EPO(22–37) fragment, bearing a dodecasaccharide carbohydrate domain. The fully elaborated biantennary N-glycan **22**, incorporating the fucose and sialic acid motifs believed to be important for in vivo EPO stability and activity, was prepared drawing heavily from convergent glycal assembly methods developed in our laboratory (Scheme 5).[29] As shown, under Lansbury aspartylation conditions **22** was successfully merged with the small peptide **23**, bearing the requisite C-terminal *ortho*-thiophenolic ester, to generate the glycopeptide **24** corresponding to the EPO(22–28) sequence.[30] Using cysteine-based NCL and OMER technology, we coupled **24** with the peptide **25**, thus generating the EPO(22–37) fragment **26** bearing the complex EPO dodecasaccharide at the native asparagine residue.

2.3. Two Auxiliary-Based Approaches to Noncysteine Glycopeptide Ligation

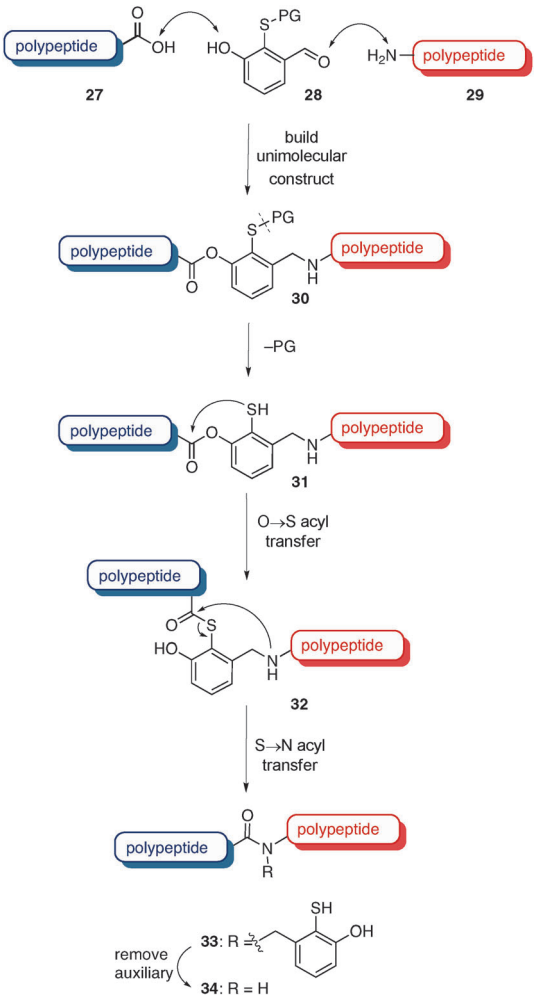
Very early on in this research program, we identified a major issue that had to be confronted in the context of our EPO total synthesis effort. The Kent NCL strategy for the merger of individual glycopeptide fragments requires that a cysteine residue be located at each site of ligation to activate the peptides for coupling. However, like many naturally occurring proteins and glycoproteins, EPO suffers from a paucity of cysteine residues. Moreover, EPO's four lonely cysteine residues are not situated, from a synthesis perspective, at convenient disconnection points along the peptide backbone. From its inception, a central focus of our EPO-directed research effort had to be on the development of a menu of broadly useful ligation capabilities, which would tap into the logic of NCL, though not require an actual cysteine at the proposed ligation sites. As described below, we have indeed developed several auxiliary-based ligation strategies, which draw mechanistic inspiration from the elegant logic of NCL. We note that prior to our studies described below the groups of Kent[31] and Dawson[32] made seminal contributions to the development of removable thiol-based auxiliaries for NCL at noncysteine sites.

2.3.1. OMER-Based Auxiliary Approach

As described above (Section 2.2), our mechanistic studies of the OMER-mediated ligation had revealed, somewhat surprisingly, that the *ortho*-thiophenolic ester undergoes ready O→S acyl transfer such that the thioester species is predominant at low pH values (2–4). In light of the evidence for the existence of a stoichiometric thioester intermediate, it occurred to us that this reactivity might be exploited for the development of an auxiliary-based cysteine-free ligation protocol. The idea, outlined in Scheme 6, is pleasingly simple.[28] First, the two peptide fragments would be mounted in a *meta* relationship on a single aromatic scaffold (**28**) by esterification (**27**) and reductive amination (**29**) methods. As shown, the aromatic auxiliary would be equipped with a protected thiol moiety in an *ortho* position relative to both peptide fragments. Upon removal of the thiol protecting group (**30**→**31**), we envisioned that a cascade sequence would initiate, commencing with O→S acyl transfer of the C-terminal polypeptide (**31**→**32**) and followed by S→N acyl

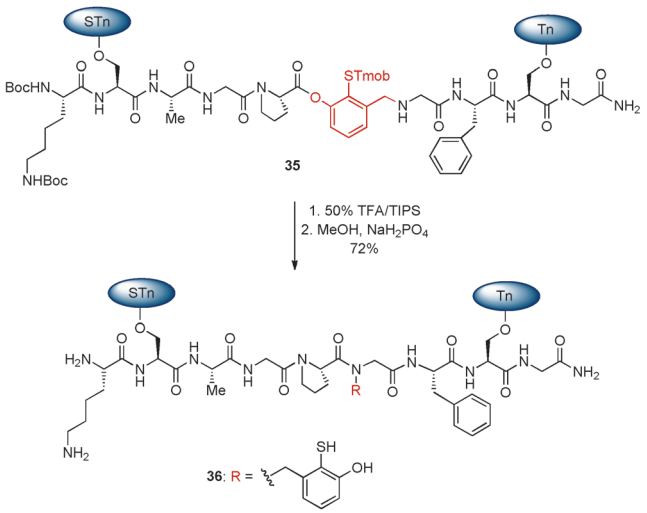


Scheme 5. Synthesis of EPO(22–37). PBS = phosphate buffered saline, TBS = *tert*-butyldimethylsilyl.



Scheme 6. OMER strategy for auxiliary based noncysteine ligation.

transfer with the amine of the second glycopeptide (**32**→**33**). Finally, removal of the auxiliary would afford the native peptide sequence (**34**).

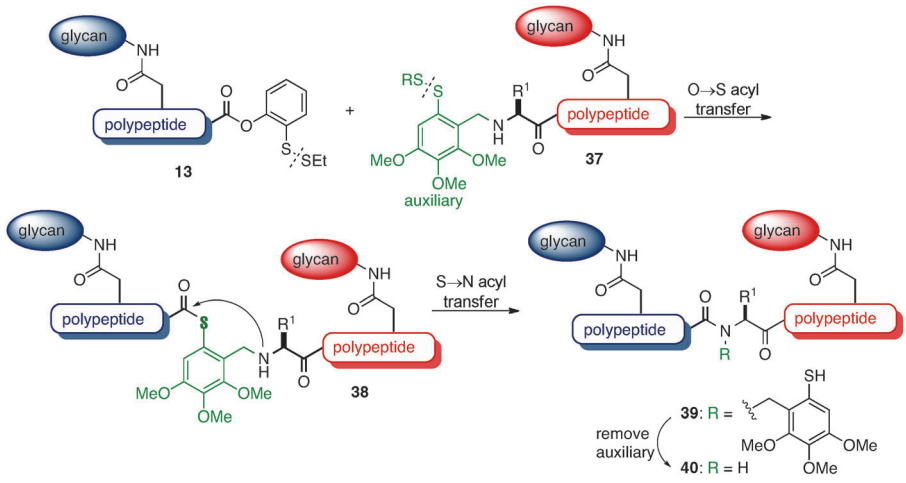


Scheme 7. Demonstration of OMER-based auxiliary protocol.

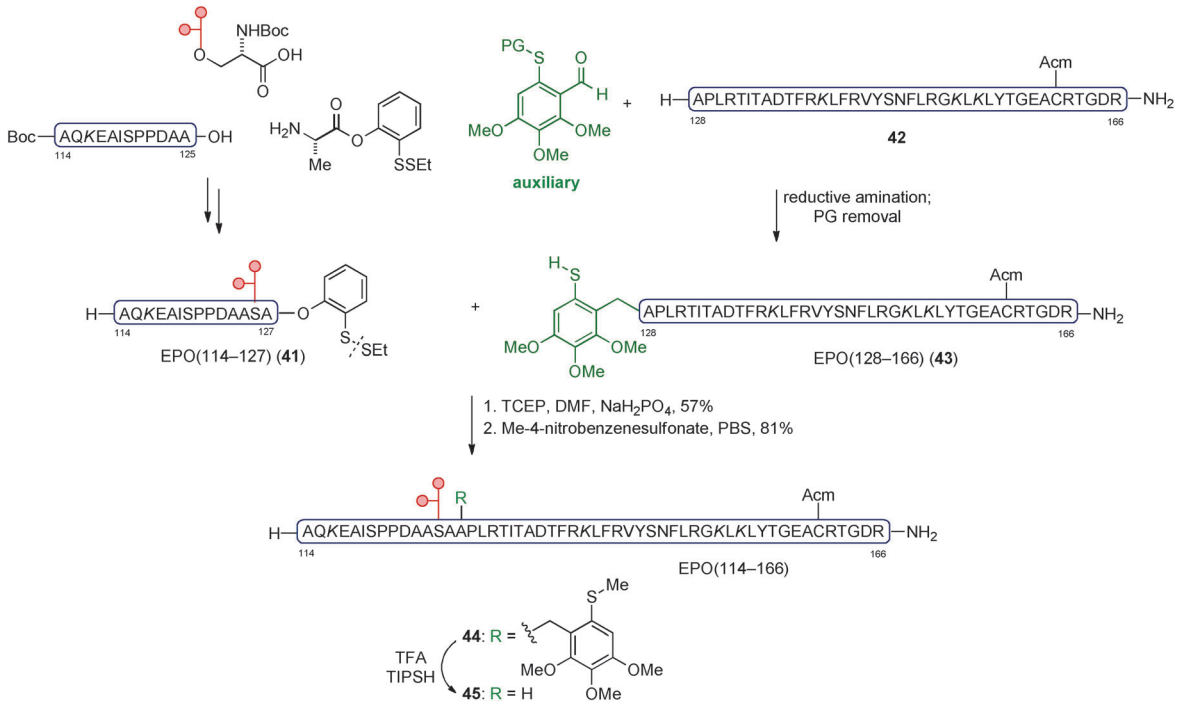
This general strategy has been reduced to practice in a number of settings including the synthesis of a bifunctional glycopeptide (Scheme 7). However, the practical application of this method to real world targets awaits, for the moment, the development of an efficient and mild protocol for removal of the auxiliary.

2.3.2. Temporary Tethered-Auxiliary Approach

We next envisioned a tethered-auxiliary approach to noncysteine ligation, wherein an N-terminal thiobenzene functionality[33] is employed to temporarily engage both glycopeptide fragments (**13** + **37**→**38**), thus bringing the reactants in sufficient proximity to undergo the key S→N acyl transfer (**38**→**39**; Scheme 8). Following a two-step auxiliary removal sequence, the native glycopeptide **40** would be in hand. The feasibility of this strategy has been demonstrated, although a practical limitation of this method arises from the requirement that a glycine or alanine be present as the C-terminal amino acid. In instances where both



Scheme 8. Tether strategy for auxiliary based noncysteine ligation.



Scheme 9. Demonstration of tether strategy: Synthesis of EPO(114–166). Boc = *tert*-butoxycarbonyl, DMF = *N*,*N*′-dimethylfirmamide, K = Lys(ivDde), TCEP = tris(2-carboxyethyl)phosphine.

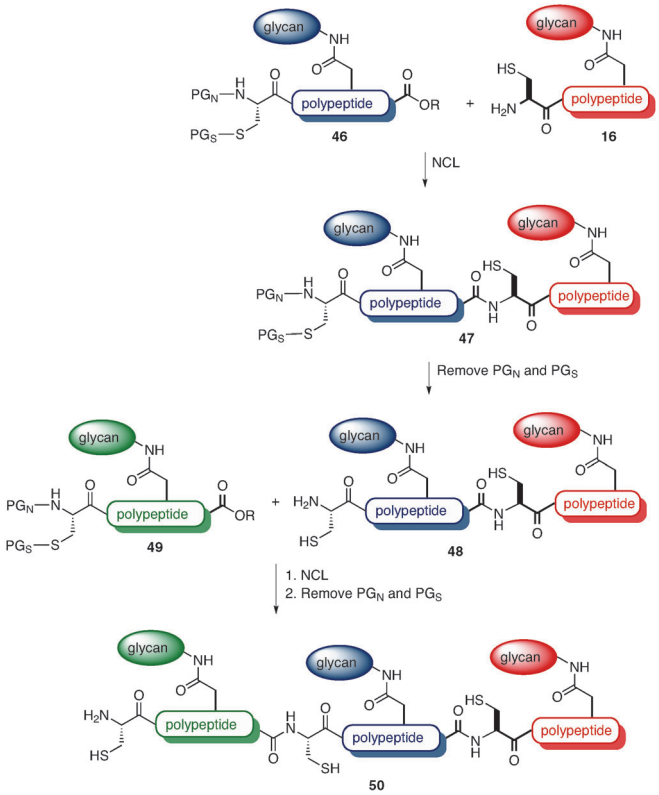
ligation site amino acids are highly branched, much diminished yields are obtained.

2.3.3. Application to EPO: Synthesis of the EPO(114–166) Fragment

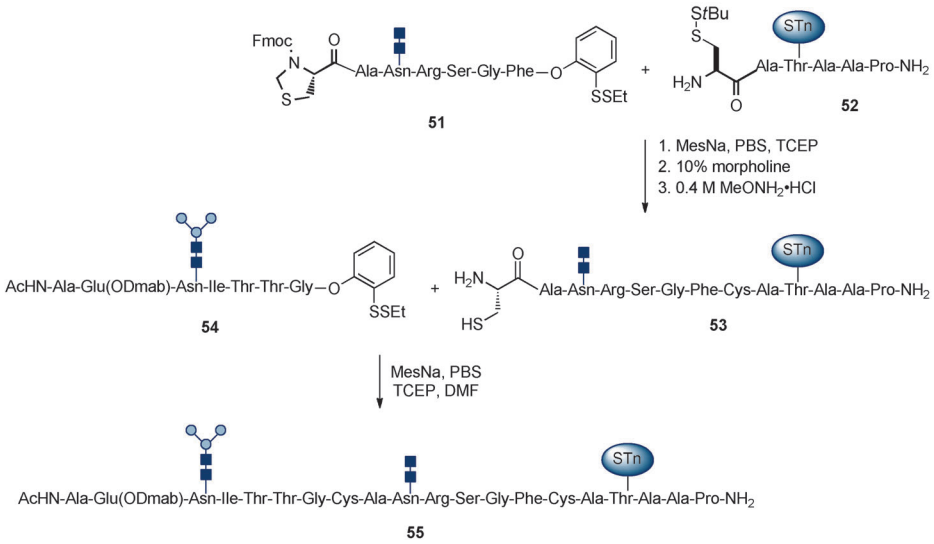
The synthetic utility of our temporary tethered-auxiliary strategy has been demonstrated in the context of a convergent synthesis of the EPO(114–166) glycopeptide fragment **45** (Scheme 9).[34] The EPO(114–127) glycopeptide **41**, bearing the O-linked glycophorin at Ser126, was synthesized according to standard methods. Separately, the peptide **42** was equipped with the thiobenzene auxiliary at its N terminus (**43**). In the event, **41** and **43** underwent auxiliary-mediated cysteine-free ligation to deliver the EPO(114–166) fragment bearing the auxiliary at Ala128. The benzyl thiol functionality was methylated to generate **44**, and upon exposure to triisopropylsilane (TIPSH) and trifluoroacetic acid (TFA) the auxiliary was released to afford the native EPO(114–166) glycopeptide **45**.

2.4. A Method for the Synthesis of Multiply Glycosylated Peptides by Reiterative Ligation

As noted above, the EPO target contains four sites of glycosylation. Our synthetic strategy envisioned the assembly of four individual glycopeptide fragments, which would be joined through a series of reiterative couplings. The challenge of sequential or reiterative peptide ligation has been studied by a number of groups with an interest in complex protein synthesis.[35] Toward this end, we endeavored to develop a method by which differentially glycosylated multifunctional peptides could be generated. The general strategy is adumbrated in Scheme 10. Two glycopeptide fragments, **46** and **16**, would be merged in the first step by OMER-facilitated NCL.



Scheme 10. Reiterative ligation strategy.



Scheme 11. Demonstration of reiterative ligation. Fmoc = 9-fluorenylmethoxycarbonyl.

The resultant bifunctional glycopeptide **47** would be equipped with a masked N-terminal cysteine residue, and after removal of the N- and S-protecting groups, the adduct **48** would be coupled with the second glycopeptide fragment **49**, thus generating a large glycopeptide bearing three different oligosaccharide domains.

In reducing this general concept to practice, we elected to employ the 1,3-thiazolidine (Thz) group as the cysteine masking functionality because of its demonstrated amenability to removal under mild reaction conditions.[36] As shown in Scheme 11, the reiterative coupling strategy outlined above was demonstrated to constitute a feasible means by which to access trifunctional peptide domains. The fragments **51**, **52**, and **54** were merged through reiterative NCL to generate the target sequence **55** in good overall yield.[37]

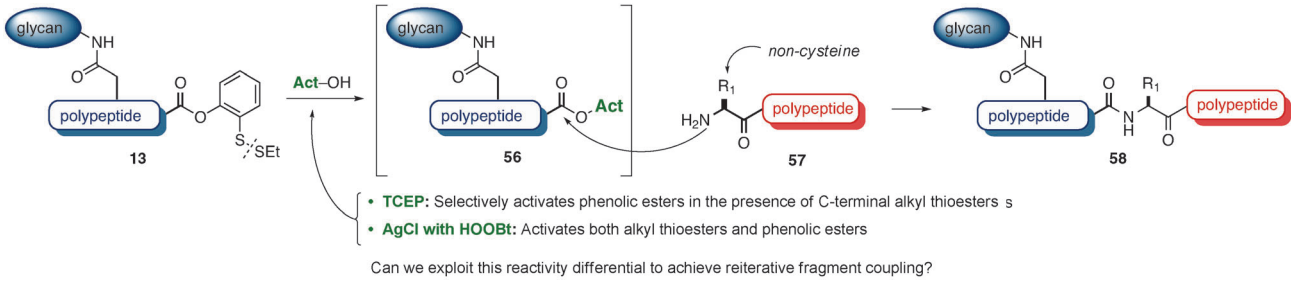
2.5. An Auxiliary-Free, Noncysteine-Based Glycopeptide Fragment Coupling

As noted above, a major goal of our laboratory has been the development of noncysteine-based ligation capabilities. As described in Section 2.3, we have developed two auxiliary-based cysteine-free methods. However, such approaches may suffer from a number of significant practical limitations. First, we encounter issues in the S→N acyl transfer when relatively hindered amino acids are located at the sites of ligation. Moreover, excision of the auxiliary from the peptide scaffold poses a challenge, particularly in the presence of vulnerable glycosidic functionality. It was in this setting that we conceived of an auxiliary-free, noncysteine-based fragment coupling approach.

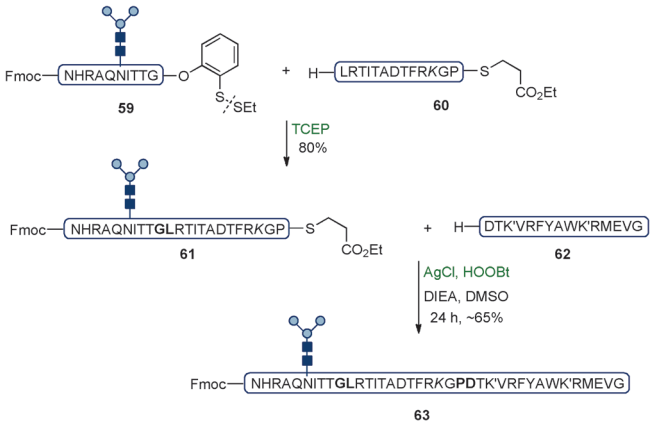
Drawing inspiration from the Blake–Aimoto silver-ion-mediated fragment condensation reaction,[38] we envisioned exploiting the unique features of the C-terminal *ortho*-thiophenolic ester **13** to generate a highly activated glycopeptide acyl donor (**56**) which might be susceptible to attack by the N terminus of a second peptide substrate (**57**; Scheme 12. The hope, of course, was that the N-terminal residue of **57** would not have to be a cysteine residue.

As shown in Scheme 12, upon investigation we did succeed in identifying two complementary sets of reaction conditions under which the hoped-for noncysteine-based fragment coupling could be achieved.[39] TCEP was found to selectively activate phenolic esters in the presence of C-terminal alkyl thioesters, while silver chloride with HOOBt is sufficiently powerful to activate both alkyl thioesters and phenolic esters.

These findings prompted us to explore whether the observed reactivity differential could be exploited to achieve



Scheme 12. Glycopeptide fragment coupling strategy.



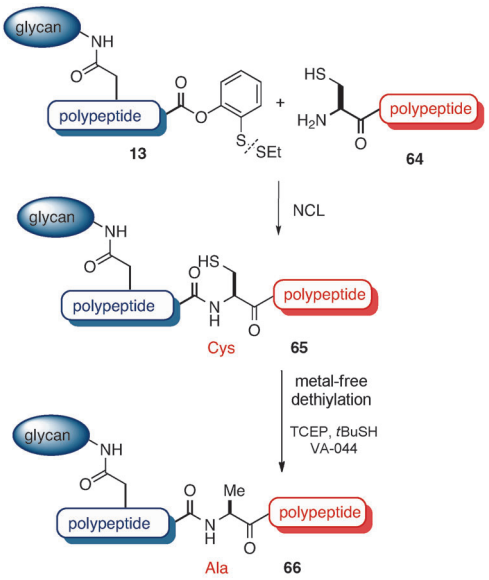
Scheme 13. Demonstration of glycopeptide fragment coupling. DIEA = diisopropylethylamine.

reiterative fragment couplings. As outlined in Scheme 13, the strategy was indeed reduced to practice. The glycopeptide **59**, bearing an *ortho*-thiophenolic ester, underwent selective TCEP-mediated fragment coupling with the peptide **60**, which is equipped with a C-terminal thioester, a group that is inert under these reaction conditions. The resultant adduct **61** readily underwent subsequent AgCl/HOOBt-mediated fragment coupling with the peptide **62** to generate the target glycopeptide fragment **63**. Despite the impressive reaction efficiency, we did find the C-terminal amino acid residue to be susceptible to epimerization in the fragment coupling step. This potential issue could be obviated through the adoption of disconnection strategies which place a proline or glycine residue at the site of ligation. As detailed below, this general strategy of noncysteine-based fragment coupling would find wide application in our first generation efforts toward the EPO glycoprotein.

2.6. A Powerful New Strategy for Noncysteine-Based Glycopeptide Ligation

In 2007, our laboratory disclosed a major methodological advance for the field of noncysteine ligation and, by extension, protein and glycoprotein synthesis.[40] In 2001, Yan and Dawson proposed, and reduced to practice, a two-step alanine ligation strategy.[41] According to their protocol, two peptide fragments are joined by cysteine-based NCL. Following ligation, the erstwhile cysteine residue is converted into alanine through exposure to Raney nickel reduction conditions. While conceptually compelling, the Dawson method suffers from significant practical limitations associated with the use of large quantities of Raney nickel for sensitive substrates. In what we and others,[4–6,42–45] perceive to be a rather useful advance, our laboratory devised a mild, free-radical-based dethiylation protocol for the conversion of cysteine into alanine residues. This method is highly tolerant of a diverse array of structural motifs commonly found in peptide and glycopeptide substrates.[40] It is well to emphasize that the core logic for this field-enhancing advance arose, not from our need for a successful synthesis, but from what would seem to be by current standards a somewhat arcane exercise in the generation and trapping of free radicals. Again, we note the critical dependence of chemical synthesis on grasp of mechanism.

We have subsequently prepared several amino acid surrogates containing appropriately placed thiol functionality. These surrogates are installed on the N-terminal fragments and undergo thiol-assisted ligation to provide adducts of the type shown in Scheme 14. Our standard metal-free dethiylation (MFD) conditions serve to readily remove the erstwhile

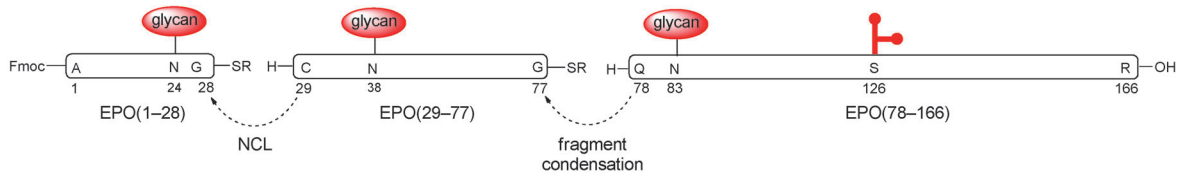


Scheme 14. Auxiliary-free noncysteine ligation: Metal-free dethiylation protocol.

thiol moieties. Through this approach, we have thus far provided the means to achieve ligation at alanine,[40] valine,[46] threonine,[47] leucine,[48] and proline[49] residues. Other laboratories have further extended the general concept of thio-amino acid ligation and subsequent desulfurization to accomplish ligations at phenylalanine,[50] lysine,[51] leucine,[52] glutamine,[53] and arginine[54] residues. With the exception of the Crich[50] phenylalanine ligation, which appeared prior to our own disclosure, each of these methods employs MFD to remove the extraneous thiol residue. As will be seen, development of this MFD protocol would prove critical to the success of the EPO endeavor.

3. A First-Generation Strategy Toward EPO

Using our upgraded problem-responsive methods described above, we set out to prepare the three glycopeptide domains which would together constitute the entire glycopeptide backbone. The general synthetic strategy, outlined in Scheme 15, called for the assembly of three glycopeptide fragments: EPO(1–28), EPO(29–77), and EPO(78–166). We



Scheme 15. First-generation synthetic strategy toward EPO.

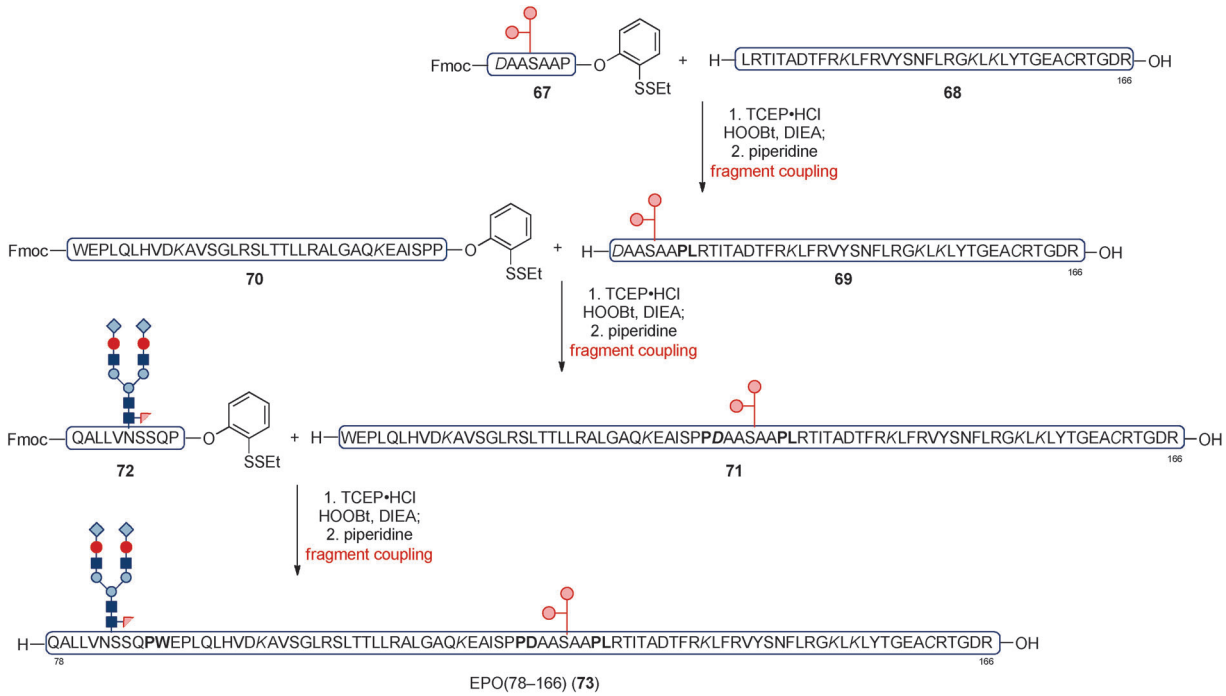
could well envision that the former two domains would be joined through cysteine-based NCL, while the merger at the 77–78 site would be accomplished through a noncysteine fragment condensation. The means by which the three individual glycopeptide fragments were assembled are outlined below.

3.1. Synthesis of the EPO(78–166) Fragment

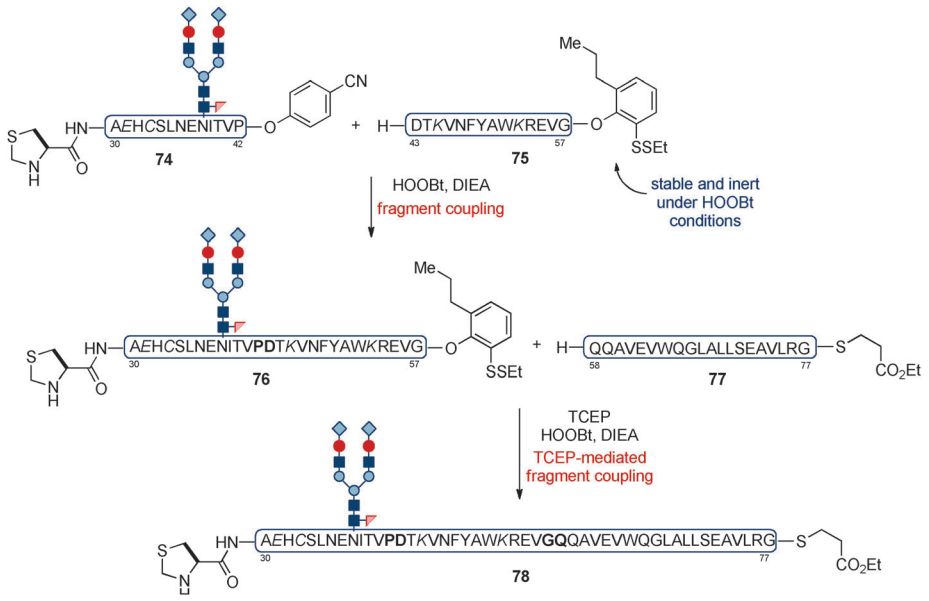
EPO(78–166), the longest of the three target subunits, incorporates both O- and N-linked glycan domains. Upon examination of the cysteine-deficient peptide sequence, we identified several well-situated proline residues, which should serve as viable sites of merger using our TCEP-mediated fragment condensation protocol (Section 2.5). We first synthesized the component side-chain protected polypeptidyl (**68** and **70**) and glycopeptidyl (**69** and **72**) substrates (Scheme 16).[55] In the event, the short glycopeptide **67**, bearing a C-terminal *ortho*-thiophenolic ester, underwent TCEP-mediated fragment coupling with the polypeptide **68** to provide the glycopeptide intermediate **69**. This intermediate was further elongated through a subsequent fragment coupling with the polypeptide **70**. In the final coupling, the glycopeptide **71**, representing the EPO(88–166) sector, was merged with the glycopeptide **72** to deliver the EPO(78–166) domain. Adoption of this fragment coupling strategy offered the crucial advantage of preserving the precious dodecasaccharide-bearing glycopeptide **72** as the limiting reagent.

3.2. Synthesis of the EPO(29–77) Fragment

The synthesis of the EPO(29–77) glycopeptide fragment proved rather challenging.[56] Our original plan called for the merger of the EPO(29–42) glycopeptide subunit with the EPO(43–77) peptide domain. However, the envisioned coupling was beset with challenges, presumably arising from poor reactivity, mismatched polarity, and severe substrate aggregation. To overcome these issues, we devised a two-step fragment coupling strategy whereby the EPO(29–42) glycopeptide fragment would be sequentially elongated in the N→C direction. The successful implementation of this strategy would require the installation of appropriately activated and differentiated C-terminal functionalities. As shown in Scheme 17, the glycopeptide **74** was equipped with a highly reactive *p*-cyanophenyl ester. The goal was to accomplish HOOBt-mediated fragment coupling with the peptide **75**,



Scheme 16. Synthesis of EPO(78–166).



Scheme 17. Synthesis of EPO(29–77).

bearing an inert *ortho*-thiophenolic ester functionality. However, upon exposure to the basic reaction conditions our standard *ortho*-thiophenolic ester underwent complete hydrolysis. Accordingly, we designed a more stable, *ortho*-substituted variant of the OMER ester with the goal of suppressing vulnerability to nucleophilic attack. In the event, the EPO(43–57) peptide **75** underwent fragment coupling with the **74** to afford the intermediate **76** in which the masked thioester had remained intact. The final goal would be that of achieving fragment coupling between **76** and the peptide **77**, bearing a C-terminal alkyl thioester functional handle. As expected (Section 2.5), under TCEP-mediated conditions, the phenolic-ester-substituted glycopeptide **76** selectively underwent fragment coupling with **77** to deliver the target EPO(29–77) fragment **78**.

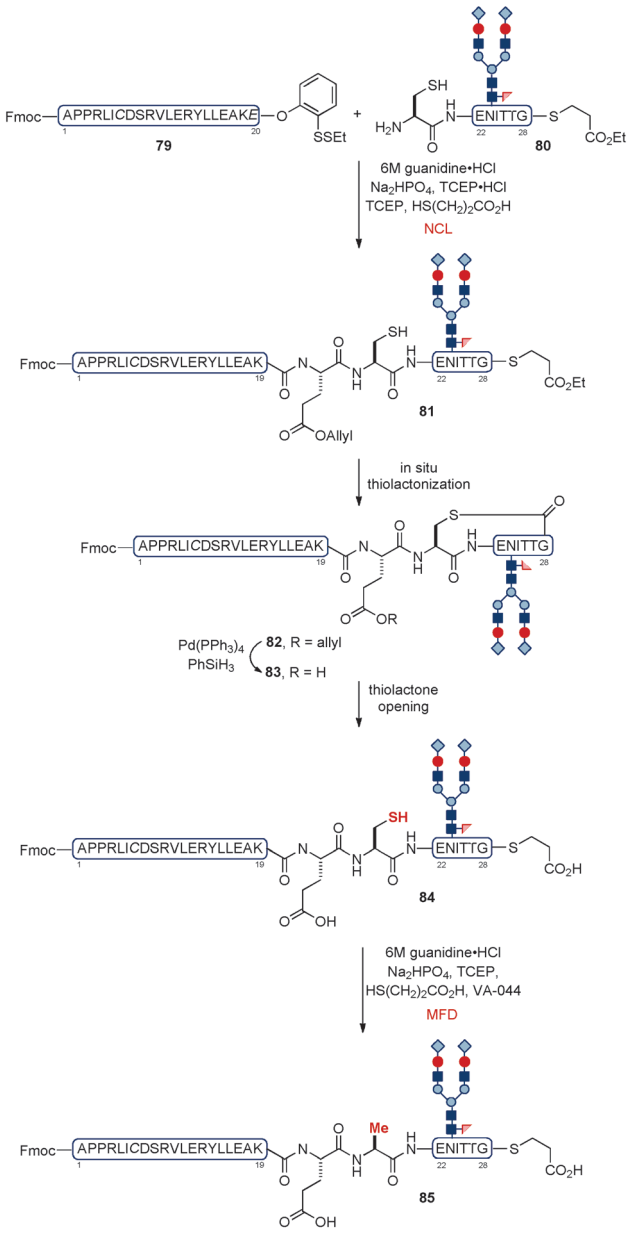
3.3. Synthesis of the EPO(1–28) Fragment

Although it is the shortest fragment, in the end the EPO(1–28) domain posed a particular synthetic challenge because of the absence of any functional cysteine, proline, or glycine residues in the peptide sequence.[57] This feature precluded the implementation of cysteine-based NCL or fragment coupling strategies. Attempts to achieve direct Lansbury aspartylation of the dodecasaccharide glycan with the full EPO(1–28) peptide segment resulted only in formation of aspartimide byproduct. To achieve the synthesis of EPO(1–28), we took recourse to a noncysteine ligation strategy. At the time of these synthetic studies, we had recently developed the highly enabling, metal-free Cys→Ala dethiylation protocol[40] described in Section 2.6. We could envision applying this method to the EPO(1–28) challenge. Along these lines, we identified Ala21 as a plausible site of disconnection. In pursuit of this strategy, we synthesized the polypeptide **79** and the shorter glycopeptide fragment **80**, bearing a temporary N-terminal Cys21 residue. In the event, **79** and **80** readily underwent NCL under standard ligation conditions to deliver the glycopeptide **81** along with the corresponding thiolactone **82** (Scheme 18). Treatment of the deprotected **83** with thiopropionic acid served to open the thiolactone, thus affording the intermediate **84**. We were now prepared to attempt the key dethiylation reaction. We were pleased to find that, upon exposure to our mild MFD conditions **84** underwent the hoped-for dethiylation to generate EPO(1–28) **85**, bearing the native Ala21 residue at the site of ligation.

With the three glycopeptide fragments in hand, we now sought to accomplish their merger according to the strategy outlined in Scheme 15. Disappointingly, however, exhaustive efforts to achieve fragment coupling of the EPO(29–77) and EPO (78–166) glycopeptides were unsuccessful and resulted in extensive aggregation and decomposition of the glycopeptide precursors (Figure 3). We ultimately recognized that a modified strategy toward EPO would be necessary.

4. A Second-Generation Strategy Toward EPO

By 2010, when we began to consider a modified EPO disconnection strategy, we had established the versatility and robustness of our noncysteine-based NCL/MFD approach in a number of complex polypeptide and glycopeptide settings.[40–48] We were thus relatively confident that this ligation technology could be successfully employed in the context of the EPO synthesis. Our second-generation strategy toward EPO envisioned the iterative merger of four glycopeptide fragments through one cysteine-based NCL and two formal alanine ligations. As shown in Scheme 19, temporary cysteine residues would be installed at positions 79 and 125. Following NCL, the erstwhile thiol functionalities would be removed by MFD to provide the native alanine residues at these sites. In



Scheme 18. Synthesis of EPO(1–28).

our opening foray, we elected to install less complex chitobiose units at each site of N-glycosylation. As will be shown, under the guidance of this general strategy we have now accomplished the first total synthesis of a wild-type, glycosylated EPO bearing chitobiose glycans.

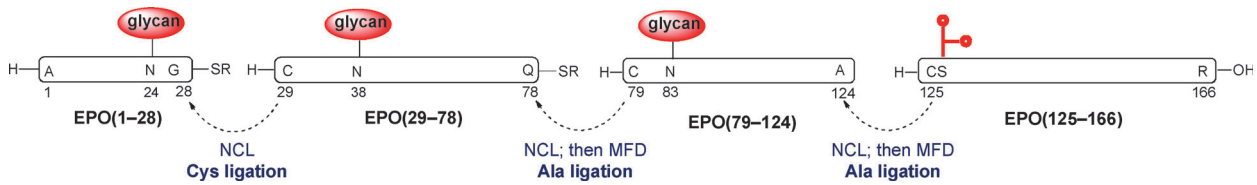
4.1. A Synthesis of the EPO(79–166) Fragment

By late 2011, we were able to describe a convergent synthesis of the EPO(79–166) fragment,[58] through a route which relies heavily on the NCL/MFD capabilities developed and optimized in our laboratory. Two long peptides (**86** and **87**) and two shorter glycopeptides (**88** and **89**) were prepared according to our standard methods (Scheme 20). Fragments **86** and **88** were merged through our pseudoproline ligation method to deliver the glycopeptide **90**, bearing a temporary thiol group on the Pro87 residue. Similarly, cysteine-based NCL of **87** and **89** delivered the glycopeptide **91**, which possesses a cysteine residue at the site of ligation. The two glycopeptide fragments, **90** and **91**, were coupled by NCL to afford the intermediate **92**, bearing three extraneous thiol groups at HS-Pro87, Cys125, and Cys128. We were pleased to find that upon exposure to our MFD conditions **92** underwent global dethiylation to generate the target EPO(79–166) fragment.

4.2. A General Strategy for Oligosaccharide Aspartylation

As noted above (Section 2.1), the preparation of N-linked glycopeptides by the Lansbury aspartylation is often severely compromised when even moderately sized polypeptide substrates are employed in the glycosylation step, because of extensive formation of aspartimide peptidyl side products.[59] The wide-reaching consequences of this significant methodological limitation, in terms of sacrifice of synthetic efficiency, can be readily appreciated. Absent the ability to directly join long polypeptide fragments with glycan coupling partners, we (and others) have generally resorted to the preparation and subsequent coupling of short glycopeptide and longer polypeptide fragments, a strategic necessity with significant ramifications for overall convergence and yield, as illustrated in Section 3.

In the context of our glycoprotein total synthesis program, we undertook to devise a solution to the longstanding problem of Lansbury aspartylation of large peptidyl substrates. Recognizing that pseudoproline dipeptides, derived from Ser or Thr, may be employed to great benefit in the solid phase peptide synthesis (SPPS) of challenging peptide sequences,[60] we took note of the fact that the generic consensus sequence for N-glycosylation in native proteins is Asn-Xaa-Ser/Thr. We wondered whether temporary installation of a pseudoproline moiety at the $(n+2)$ position relative to the Asp residue might, in some way, serve to suppress aspartimide formation, perhaps by altering the local structural or electronic character of the peptide around the



Scheme 19. Second-generation synthetic strategy toward EPO.

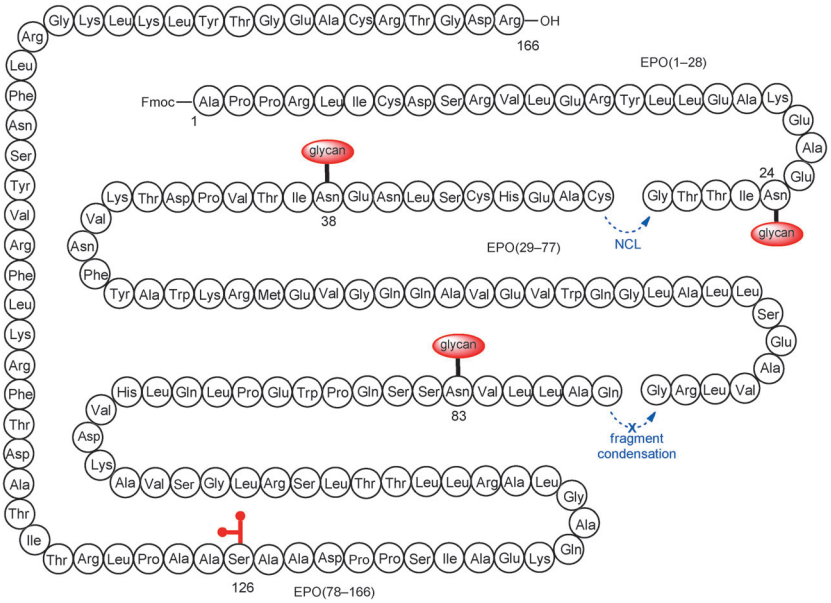
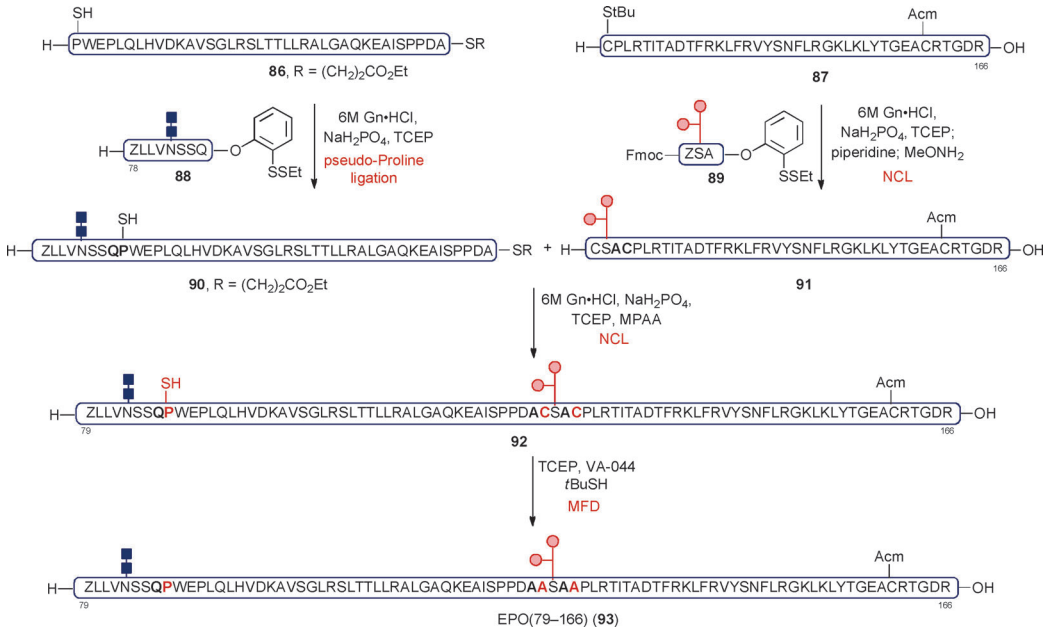


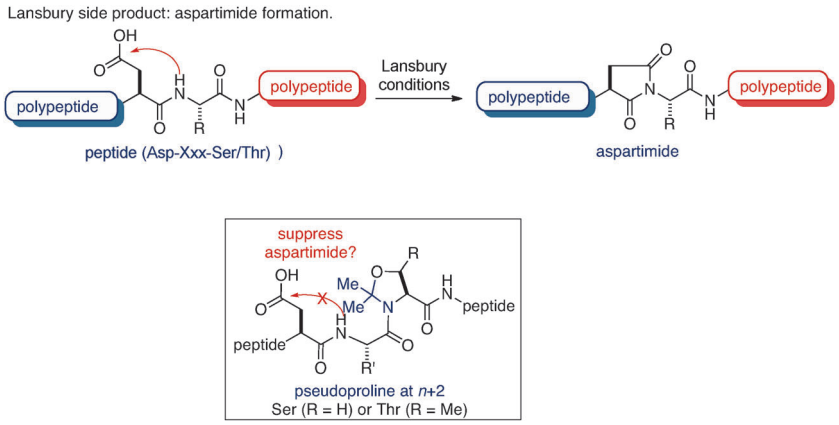
Figure 3. First-generation strategy toward EPO.



Scheme 20. Synthesis of EPO(79–166).

aspartylation site. Because a Ser or Thr residue is universally located at the $(n+2)$ position, such a strategy, if successful, would represent a general solution to the very serious problem of efficient peptide–glycan coupling.

In 2012, we accomplished the implementation of the pseudoproline-based aspartylation concept encompassed in Scheme 21.[61] Actually, the method was initially devised to meet the challenge associated with the synthesis of the FSH α and β subunits.[20] Under our optimized one-flask aspartylation/deprotection conditions, fully protected polypeptide sequences of significant length are subjected to Lansbury aspartylation with glycan domains to afford, after deprotection, fully elaborated glycopeptide fragments. As outlined below, the development of this powerful synthetic capability has now enabled the attainment of a remarkably rapid and convergent total synthesis of the EPO glycoprotein. Interestingly, the group of Unverzagt independently conceived of the same solution to the Lansbury aspartylation aspartimide formation problem. Their elegant disclosure, describing the effects of installation of a similar pseudoproline dipeptide substructural motif at the $(n+2)$ position in solid-phase glycopeptide synthesis, appeared on *Angewandte Chemie* (Early View) shortly prior to the publication of our own work.[62]



Scheme 21. Proposed solution to the problem of aspartimide formation in Lansbury aspartylation.

4.3. Highly Convergent Synthesis of the EPO(79–124), EPO(29–78), and EPO(1–28) Fragments

Each N-linked glycopeptide fragment—EPO(79–124), EPO(29–78), and EPO(1–28)—was prepared according to the one-flask aspartylation/deprotection strategy described above.[16] The protected peptide precursors (**94**, **96**, and **98**), incorporating pseudoproline dipeptide motifs, were synthesized through SPPS (Scheme 22). Under Lansbury conditions, each peptide readily underwent coupling with chitobiose glycan. Subsequent addition of a TFA cocktail (TFA/TIPSH/H_2O/phenol) served to unmask the pseudoproline functionalities and remove the peptide-protecting groups, thereby delivering the target glycopeptide fragments, **95**, **97**, and **99**, bearing the requisite cysteine-protecting groups and C-terminal functionalization. The fourth EPO fragment, encompassing the EPO(125–166) sector, was prepared according to our previously described route[58] (**91**; Scheme 20). The remarkable synthetic advantage offered by our newly developed aspartylation method can be well appreciated in this setting. Each of the three N-linked glycopeptide fragments was prepared in a single step from an SPPS-derived precursor.

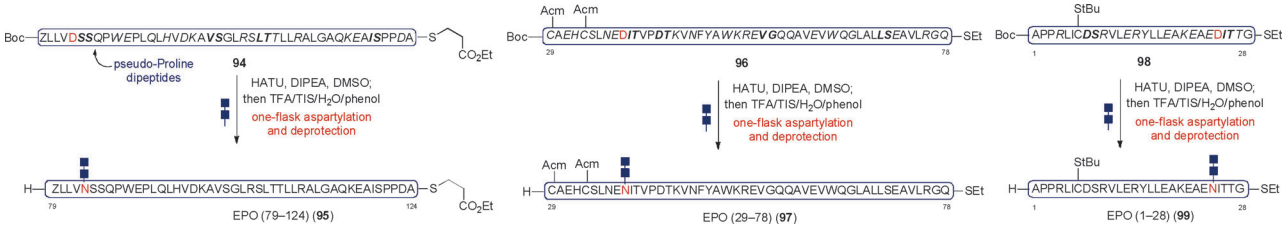
4.4. First Total Synthesis of the EPO Primary Structure

With the four EPO glycopeptide fragments in hand, we now sought to explore the coupling strategy outlined in Scheme 19. The glycopeptides **95** and **91** were merged under standard NCL conditions to provide the intermediate **100**, which encompasses the EPO(79–166) domain (Scheme 23). This construct underwent a second cysteine-based NCL with **97** to deliver the EPO(29–166) fragment **101**. At this stage, exposure to MFD conditions served to convert the three extraneous cysteines into the requisite alanine residues at positions 79, 125, and 128. Upon removal of the cysteine Acm protecting groups,[63] the glycopeptide **103** was obtained. In the final step, cysteine-based NCL with EPO(1–28) (**99**) afforded the EPO(1–166) primary structure **104**, which possesses the wild-type amino acid sequence and all four native sites of glycosylation.

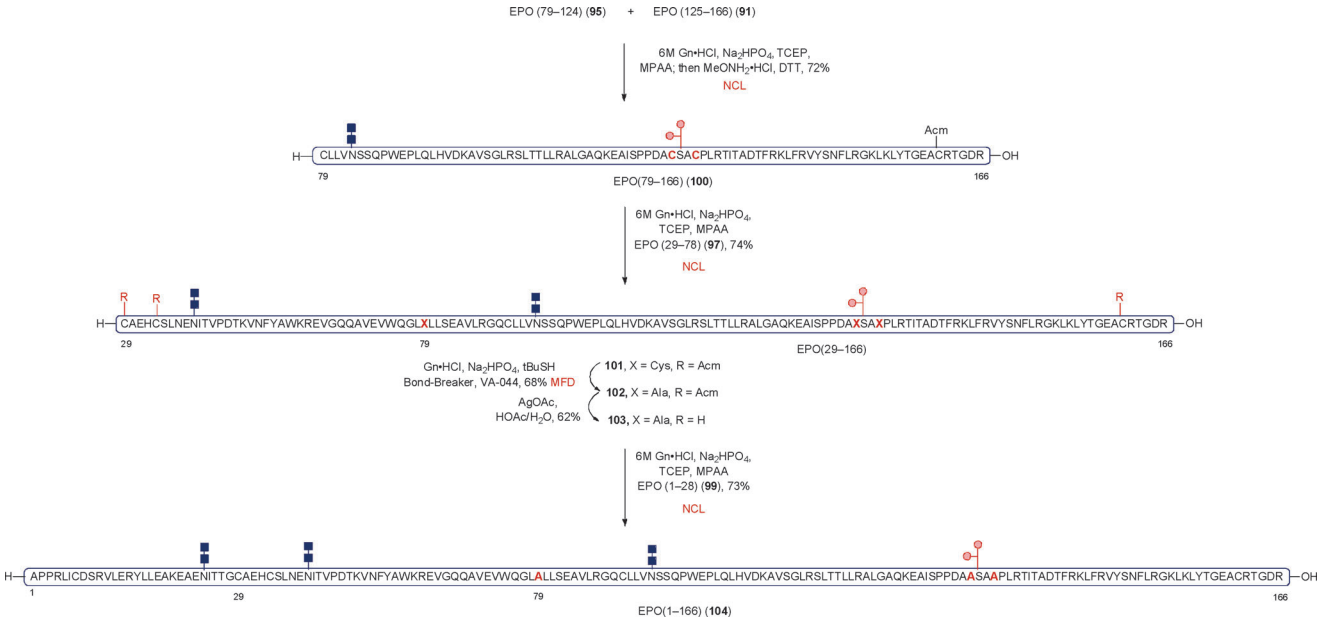
4.5. Total Synthesis of Folded EPO Glycoprotein

In a parallel effort, we explored an alternative, kinetically controlled ligation[64] (KCL)-based route to EPO, with the goal of further optimizing synthetic convergence.[16] The idea was to accomplish a one-flask merger of the three component EPO fragments. We reasoned that if the site of the first disconnection were shifted from Gly28-Cys29 to Cys29-Ala30, it should be possible to accomplish a series of kinetically controlled cysteine-based ligations (using a temporary cysteine in place of the alanine at position 30) to deliver the full glycopeptide sequence bearing non-native cysteine residues at positions 30, 79, 125, and 128. A single global MFD step followed by Acm removal, should reveal the EPO primary structure.

This strategy proved to be highly effective. The modified glycopeptide fragments **105** and **106**, bearing differentiated C-terminal thioester functionalities, were prepared through recourse to the one-flask aspartylation/deprotection method



Scheme 22. Demonstration of aspartylation protocol: Synthesis of EPO glycopeptide fragments.



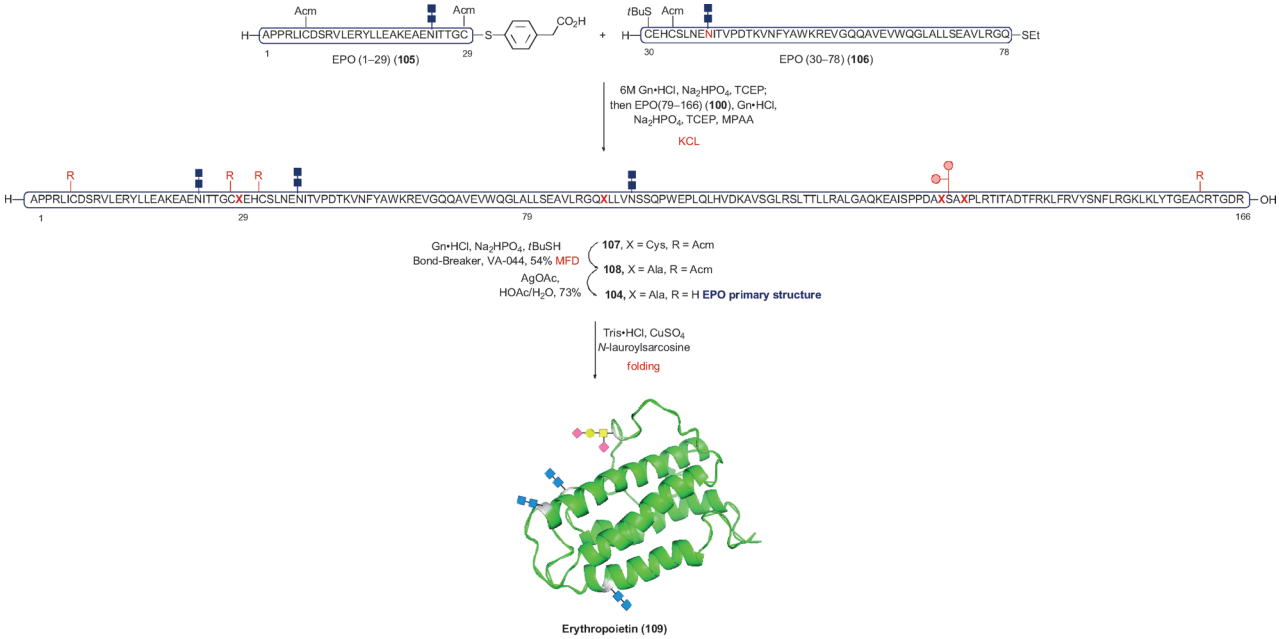
Scheme 23. Total synthesis of the EPO primary structure. DTT = dithiothreitol, MPAA = 4-mercaptophenylacetic acid.

described above.[61] KCL of **105** and **106**, followed by in situ activation of the Gln alkylthioester in the presence of **100** afforded the target glycopeptide **107**. The crude mixture was subjected to MFD conditions to generate **108**, and subsequent deprotection delivered the EPO(1–166) primary structure (**104**). The HPLC retention time and mass spectral data obtained from both routes (Scheme 23 and 24) were identical.

We were now poised to attempt the folding of a fully synthetic, suitably glycosylated EPO primary structure according to the literature reported protocol reported for CHO cell-derived inhomogeneous EPO. In the event, following exposure of **104** to $CuSO_4$ oxidant and N-lauroylsarcosine additive, the fully folded homogeneous EPO glycoprotein was obtained. The reader can well imagine our delight (and relief) when top-down mass spectroscopy analysis confirmed formation of the folded glycoprotein.

4.6. Total Synthesis of Nonglycosylated Folded EPO Protein

To examine the effects of protein glycosylation on the physical properties of EPO, we adapted the synthetic sequence outlined in Section 4.4 to gain access to nonglycosylated EPO protein. Our previously described efforts to



Scheme 24. Total synthesis of folded EPO.

Scheme 25. Total synthesis of folded nonglyco EPO.

accomplish the total synthesis of nonglycosylated EPO had been beset by complications arising from the very low solubility of the fully assembled, nonglycosylated peptide chain.[65] However, the route employed in Section 4.4 offered a significant advantage, in that the assembly of the full peptide sequence represents the last step of the synthesis prior to protein folding (Scheme 25). Our efforts to accomplish the synthesis and folding of nonglycosylated EPO protein were indeed successful, although aggregation and solubility issues severely complicated the acquisition of high quality mass spectrometry data. These studies served to confirm the admirable intuition of CHO in glycosylating its nascent EPO, even in the form of a horrific mixture of glycoforms.

4.7. Fully Synthetic Folded EPO Glycoprotein Possesses Erythropoietic Activity

Finally, we could evaluate the erythropoietic activity of unfolded EPO primary sequence (**104**), as well as unglycosylated EPO protein (**119**), and folded EPO glycoprotein (**109**). Although unfolded EPO(1–166) failed to exhibit detectable activity, folded EPO glycoprotein (**109**) and folded nonglycoEPO (**119**) demonstrated measurable biological activity. In in vitro cell proliferation studies, both **109** and **119** promoted formation of erythroid colonies from progenitor CD34+ cells. The glycosylated EPO demonstrated much enhanced activity in comparison with the nonglycosylated EPO, perhaps because of the poor stability of the nonglycosylated EPO (Figure 4).

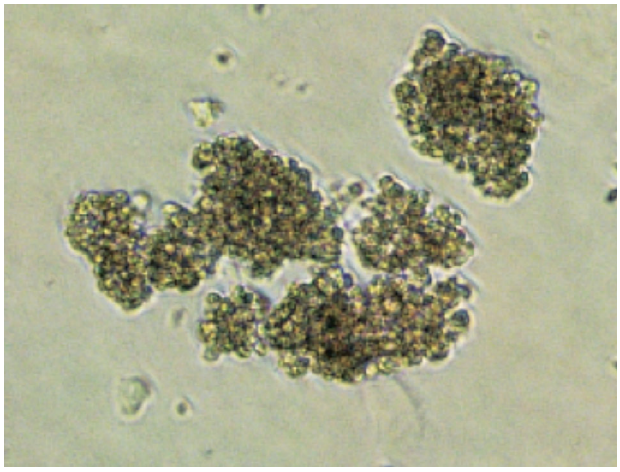


Figure 4. The image of human burst forming unit-erythrocyte (BFU-E) colony. 60 ng mL^{-1} synthetic folded EPO (**109**) and 20 ng mL^{-1} rhKL stimulate purified cord blood CD34 cells to form BFU-E colony after 2 weeks

5. Conclusion

We have presented an account of the trials, tribulations, and progress in completing our ten-year journey to a functional EPO. Happily, we were able to generate the full primary structure without any accommodations for the sake of simplifying the problem. Moreover, we successfully introduced glycosylation at each of the native sites. It was shown that the prefolded EPO did not exhibit biological function, while the folded counterpart did indeed induce erythropoiesis. Hopefully, the stage is now set to acquire even greater insights into the complex structure–activity relationship (SAR) issues governing the effects of glycosylation on the performance of EPO. To do ultimate justice to this

ambitious goal, it will be necessary to generate a range of N-glycosylated congeners of varying complexity at the native asparagine sites. In this way, it could be possible to evaluate the effects of glycosylation on three most interesting questions: First, we hope to be able to provide more reliable information on the impact of glycosylation on EPO folding. Second, we plan to assess the relationship between EPO glycosylation and erythropoiesis activity. Finally, we will assess the consequences of glycosylation on in vivo performance, even in higher organisms. Ultimately, to accomplish such goals, it will be necessary to expand our chemistry to more ambitious glycodomains. Progress in this area is ongoing and we are confident that the chemistry charted above will be extendable to more complex settings. In the meantime, even as these challenges remain to be fully overcome, and as progress goes forward, it is perhaps fair to say that much has already been learned.

All this said, we chance to transition from the relatively secure terrain of reportage to the rather riskier dimension of prediction. The exercise described above has served to enhance our confidence that chemical synthesis is on the cusp of being able to bring into being highly complex structures hitherto perceived to be available only through strictly biological means. At this stage, it is really is too early to become seriously engrossed as to the commercializibility of the chemistry enabled synthesis, relative to synthesis through traditional biological means. First, the generality and limitations of these EPO driven advances must be established. Application to other high-performing biologics (cytokines, vaccines) will help to define the limits of the emerging capabilities. Moreover, there will surely emerge problems in the synthesis of biologics which are best solved by combining the optimized skill sets of the two fields in a synergistic way. For instance, enzymatically mediated synthesis of oligosaccharides may be interfaced with aspartylation and MFD-driven NCL to reach homogeneous, performance-optimized glycopolypeptides. It is only when these and other capabilities are better sorted out that sensible predictions as to commercializibility can emerge.

We close with the thought that the cause of optimization of the synergy of the two bedrock sciences we now call biology and chemistry would be well served by sensitive attention to issues of cross-cultural language. It may be well for each of the core disciplines to be increasingly responsive to the particular passions of the other. We are hopeful that this account of our travails and triumphs in pursuing our obsession with EPO[66] will serve the very important goal of cross-science collegiality.

Support for this research was provided by the National Institute of Health (HL25848 to S.J.D.).

Received: February 26, 2013
Published online: June 17, 2013

[1] For the first example of NCL in glycopeptide synthesis, see: Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman, C. R. Bertozzi, *J. Am. Chem. Soc.* **1999**, *121*, 11684.

[2] For the synthesis of alymphotactin, a glycosylated chemokine with a C-terminal mucin-like domain, see: L. A. Marcaurelle, L. S. Mizoue, J. Wilken, L. Oldham, S. B. H. Kent, T. M. Handel, C. R. Bertozzi, *Chem. Eur. J.* **2001**, *7*, 1129.

[3] For the synthesis of the 76-amino acid chemokine monocyte chemotactic protein-3, see: N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson, Y. Kajihara, *J. Am. Chem. Soc.* **2008**, *130*, 501.

[4] For the synthesis of ribonuclease C, see: a) C. Piontek, P. Ring, O. Harjes, C. Heinlein, S. Mezzato, N. Lombana, C. Pöhner, M. Püttner, D. V. Silva, A. Martin, F. X. Schmid, C. Unverzagt, *Angew. Chem.* **2009**, *121*, 1968; *Angew. Chem. Int. Ed.* **2009**, *48*, 1936; b) C. Piontek, D. V. Silva, C. Heinlein, C. Pöhner, S. Mezzato, P. Ring, A. Martin, F. X. Schmid, C. Unverzagt, *Angew. Chem.* **2009**, *121*, 1974; *Angew. Chem. Int. Ed.* **2009**, *48*, 1941.

[5] For the synthesis of antifreeze glycoproteins, see: B. L. Wilkinson, R. S. Stone, C. J. Capicciotti, M. Thaysen-Andersen, J. M. Matthews, N. H. Packer, R. N. Ben, R. J. Payne, *Angew. Chem.* **2012**, *124*, 3666; *Angew. Chem. Int. Ed.* **2012**, *51*, 3606.

[6] For the synthesis of interferon-β, see: I. Sakamoto, K. Tezuka, K. Fukae, K. Ishii, K. Taduru, M. Maeda, M. Ouchi, K. Yoshida, Y. Nambu, J. Igarashi, N. Hayashi, T. Tsuji, Y. Kajihara, *J. Am. Chem. Soc.* **2012**, *134*, 5428.

[7] For general reviews on EPO, see: a) S. Elliott, M. A. Foote, G. Molineux, *Erythropoietins, Erythropoietic Factors, and Erythropoiesis, 2nd ed.*, Birkhäuser, Boston, **2009**; b) A. J. Sytkowski, *Erythropoietin*, Wiley-VCH, Weinheim, **2004**; c) J. Szenajch, G. Wcislo, J.-Y. Jeong, C. Szczylik, L. Feldman, *Biochim. Biophys. Acta Rev. Cancer* **2010**, *1806*, 82.

[8] a) M. Higuchi, M. Oh-eda, H. Kuboniwa, K. Tomonoh, Y. Shimonaka, *J. Biol. Chem.* **1992**, *267*, 7703; b) J. C. Egrie, J. R. Grant, D. K. Gillies, K. H. Aoki, T. W. Strickland, *Glycoconjugate J.* **1993**, *10*, 263.

[9] a) R. Kornfeld, S. Kornfeld, *Annu. Rev. Biochem.* **1985**, *54*, 631; b) J. Roth, *Chem. Rev.* **2002**, *102*, 285; c) P. M. Rudd, R. A. Dwek, *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, 1.

[10] D. Macmillan, R. M. Bill, K. A. Sage, D. Fern, S. L. Flitsch, *Chem. Biol.* **2001**, *8*, 133.

[11] G. G. Kochendoerfer, S.-Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent, J. A. Bradburne, *Science* **2003**, *299*, 884.

[12] S. Liu, B. L. Pentelute, S. B. H. Kent, *Angew. Chem.* **2012**, *124*, 1017; *Angew. Chem. Int. Ed.* **2012**, *51*, 993.

[13] J. P. Richardson, D. Macmillan, *Org. Biomol. Chem.* **2008**, *6*, 3977.

[14] K. Hirano, D. Macmillan, K. Tezuka, T. Tsuji, Y. Kajihara, *Angew. Chem.* **2009**, *121*, 9721; *Angew. Chem. Int. Ed.* **2009**, *48*, 9557.

[15] M. Murakami, R. Okamoto, M. Izumi, Y. Kajihara, *Angew. Chem.* **2012**, *124*, 3627; *Angew. Chem. Int. Ed.* **2012**, *51*, 3567.

[16] P. Wang, S. Dong, J. A. Brailsford, K. Iyer, S. D. Townsend, Q. Zhang, R. C. Hendrickson, J. H. Shieh, M. A. S. Moore, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 11744; *Angew. Chem. Int. Ed.* **2012**, *51*, 11576.

[17] C. Kan, S. J. Danishefsky, *Tetrahedron* **2009**, *65*, 9047.

[18] a) S. Shang, Z. Tan, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4297; b) S. Dong, S. Shang, J. Li, Z. Tan, T. Dean, A. Maeda, T. J. Gardella, S. J. Danishefsky, *J. Am. Chem. Soc.* **2012**, *134*, 15122.

[19] J. Li, S. Dong, S. D. Townsend, T. Dean, T. J. Gardella, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 12429; *Angew. Chem. Int. Ed.* **2012**, *51*, 12263.

[20] a) P. Nagorny, N. Sane, B. Fasching, B. Aussedat, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 999; *Angew. Chem. Int. Ed.* **2012**, *51*, 975; b) B. Aussedat, B. Fasching, E. Johnston, N. Sane, P. Nagorny, S. J. Danishefsky, *J. Am. Chem. Soc.* **2012**, *134*, 3532.
[21] S. J. Danishefsky, M. T. Bilodeau, *Angew. Chem.* **1996**, *108*, 1482; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1380.
[22] J. S. Miller, V. Y. Dudkin, G. J. Lyon, T. W. Muir, S. J. Danishefsky, *Angew. Chem.* **2003**, *115*, 447; *Angew. Chem. Int. Ed.* **2003**, *42*, 431.
[23] L. M. Likhosherstov, O. S. Novikova, V. A. Derevitskaja, N. K. Kochetkov, *Carbohydr. Res.* **1986**, *146*, C1.
[24] a) S. T. Anisfeld, P. T. Lansbury, *J. Org. Chem.* **1990**, *55*, 5560; b) S. T. Cohen-Anisfeld, P. T. Lansbury, *J. Am. Chem. Soc.* **1993**, *115*, 10531.
[25] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776.
[26] J. D. Warren, J. S. Miller, S. J. Keding, S. J. Danishefsky, *J. Am. Chem. Soc.* **2004**, *126*, 6576.
[27] J. Chen, J. D. Warren, B. Wu, G. Chen, Q. Wan, S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 1969.
[28] G. Chen, J. D. Warren, J. Chen, B. Wu, Q. Wan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, *128*, 7460.
[29] B. Wu, Z. Hua, J. D. Warren, K. Ranganathan, Q. Wan, G. Chen, Z. Tan, J. Chen, A. Endo, S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 5577.
[30] B. Wu, Z. Tan, G. Chen, J. Chen, Z. Hua, Q. Wan, K. Ranganathan, S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 8009.
[31] a) L. E. Canne, S. J. Bark, S. B. H. Kent, *J. Am. Chem. Soc.* **1996**, *118*, 5891; b) D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554.
[32] J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642.
[33] B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua, S. J. Danishefsky, *Angew. Chem.* **2006**, *118*, 4222; *Angew. Chem. Int. Ed.* **2006**, *45*, 4116.
[34] J. Chen, G. Chen, B. Wu, Q. Wan, Z. Tan, Z. Hua, S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 8013.
[35] For a recent review, see: L. Raibaut, N. Ollivier, O. Melnyk, *Chem. Soc. Rev.* **2012**, *41*, 7001.
[36] D. Bang, S. B. Kent, *Angew. Chem.* **2004**, *116*, 2588; *Angew. Chem. Int. Ed.* **2004**, *43*, 2534.
[37] B. Wu, J. D. Warren, J. Chen, G. Chen, Z. Hua, S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 5219.
[38] a) J. Blake, *Int. J. Peptide Protein Res.* **1981**, *17*, 273–274; b) S. Aimoto, N. Mizoguchi, H. Hojo, S. Yoshimura, *Bull. Chem. Soc. Jpn.* **1989**, *62*, 524–531; c) S. Aimoto, *Biopolymers* **1999**, *51*, 247–265.
[39] G. Chen, Q. Wan, Z. Tan, C. Kan, Z. Hua, K. Ranganathan, S. J. Danishefsky, *Angew. Chem.* **2007**, *119*, 7527; *Angew. Chem. Int. Ed.* **2007**, *46*, 7383.
[40] Q. Wan, S. J. Danishefsky, *Angew. Chem.* **2007**, *119*, 9408; *Angew. Chem. Int. Ed.* **2007**, *46*, 9248.
[41] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526.
[42] C. Haase, H. Rohde, O. Seitz, *Angew. Chem.* **2008**, *120*, 6912; *Angew. Chem. Int. Ed.* **2008**, *47*, 6807.
[43] K. S. Ajish Kumar, M. Haj-Yahya, D. Olschewski, H. A. Lashuel, A. Brik, *Angew. Chem.* **2009**, *121*, 8234; *Angew. Chem. Int. Ed.* **2009**, *48*, 8090.
[44] S. N. Bavikar, L. Spasser, M. Haj-Yahya, S. V. Karthikeyan, T. Moyal, K. S. A. Kumar, A. Brik, *Angew. Chem.* **2012**, *124*, 782; *Angew. Chem. Int. Ed.* **2012**, *51*, 758.
[45] B. Fierz, S. Kilic, A. R. Hieb, K. Luger, T. W. Muir, *J. Am. Chem. Soc.* **2012**, *134*, 19548.
[46] J. Chen, Q. Wan, Y. Yuan, J. L. Zhu, S. J. Danishefsky, *Angew. Chem.* **2008**, *120*, 8649; *Angew. Chem. Int. Ed.* **2008**, *47*, 8521.
[47] J. Chen, P. Wang, J. Zhu, Q. Wan, S. J. Danishefsky, *Tetrahedron* **2009**, *66*, 2277.
[48] Z. Tan, S. Shang, S. J. Danishefsky, *Angew. Chem.* **2010**, *122*, 9690; *Angew. Chem. Int. Ed.* **2010**, *49*, 9500.
[49] S. Shang, Z. Tan, S. Dong, S. J. Danishefsky, *J. Am. Chem. Soc.* **2011**, *133*, 10784.
[50] D. Crich, A. Banerjee, *J. Am. Chem. Soc.* **2007**, *129*, 10064.
[51] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *J. Am. Chem. Soc.* **2009**, *131*, 13592.
[52] Z. Harpaz, P. Siman, K. S. A. Kumar, A. Brik, *ChemBioChem* **2010**, *11*, 1232.
[53] P. Siman, S. V. Karthikeyan, A. Brik, *Org. Lett.* **2012**, *14*, 1520.
[54] L. R. Malins, K. M. Cergol, R. J. Payne, *ChemBioChem* **2013**, *14*, 559.
[55] Z. Tan, S. Shang, T. Halkina, Y. Yuan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5424.
[56] Y. Yuan, J. Chen, Q. Wan, Z. Tan, G. Chen, C. Kan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5432.
[57] C. Kan, J. D. Trzupek, B. Wu, Q. Wan, G. Chen, Z. Tan, Y. Yuan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5438.
[58] S. Dong, S. Shang, Z. Tan, S. J. Danishefsky, *Isr. J. Chem.* **2011**, *51*, 968.
[59] a) M. Bodanszky, J. C. Tolle, S. S. Deshmane, A. Bodanszky, *Int. J. Pept. Protein Res.* **1978**, *12*, 57; b) M. Bodanszky, G. F. Sigler, A. Bodanszky, *J. Am. Chem. Soc.* **1973**, *95*, 2352.
[60] F. García-Martín, P. White, R. Steinauer, S. Côté, J. Tulla-Puche, F. Albericio, *Biopolymers* **2006**, *84*, 566.
[61] P. Wang, B. Aussedat, Y. Vohra, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 11739; *Angew. Chem. Int. Ed.* **2012**, *51*, 11571.
[62] V. Ullmann, M. Radisch, I. Boos, J. Freund, C. Pohner, S. Schwarzinger, C. Unverzagt, *Angew. Chem.* **2012**, *124*, 11734; *Angew. Chem. Int. Ed.* **2012**, *51*, 11566.
[63] D. Bang, N. Chopra, S. B. Kent, *J. Am. Chem. Soc.* **2004**, *126*, 1377.
[64] a) D. Bang, B. Pentelute, S. B. H. Kent, *Angew. Chem.* **2006**, *118*, 4089; *Angew. Chem. Int. Ed.* **2006**, *45*, 3985; For mechanistic studies, see: b) E. C. B. Johnson, S. B. H. Kent, *J. Am. Chem. Soc.* **2006**, *128*, 6640.
[65] J. A. Brailsford, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7196.
[66] Of course, while all this had been in progress, there was a corresponding infatuation with a small-molecule family known as the epothilones. Confusion as to these two ventures was minimized by use of descriptors (big EPO and little Epo). This led, eventually, to the non-natural small molecules fludelone and isofludelone. The latter is currently in human clinical trials in oncology settings. See: a) A. Rivkin, T.-C. Chou, S. J. Danishefsky, *Angew. Chem.* **2005**, *117*, 2898; *Angew. Chem. Int. Ed.* **2005**, *44*, 2838; b) T.-C. Chou, X. Zhang, Z.-Y. Zhong, Y. Li, L. Feng, S. Eng, D. R. Myles, R. Johnson, N. Wu, Y. I. Yin, R. M. Wilson, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13157.